

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:40:04 ; Search time 36 Seconds
(without alignments)
1311.767 Million cell updates/sec

Title: US-09-589-870B-2

Perfect score: 936
Sequence: 1 MKRIVAAIAVSLTVSITA.....IDAKKAGVNGNPLDAVQO 183

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

SPREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp Vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvivirus:*
- 16: sp_bacteriophage:*
- 17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query #	Match Length	DB ID	Description
1	168.5	18.0	179	2	08KXW2
2	145.5	15.5	152	13	096SH4
3	114	12.2	529	5	025058
4	110.5	11.8	892	2	09KX38
5	105.5	11.3	1001	3	005164
6	104	11.1	595	2	09S0T7
7	102	10.9	1310	16	08Y4J4
8	102	10.9	1371	16	08XQ42
9	102	10.8	1374	16	08YJM4
10	101	10.8	595	2	006351
11	100	10.7	560	2	09S4M4
12	100	10.7	565	2	09S4M4
13	99.5	10.6	1477	2	048028
14	99.5	10.6	2468	16	0912M3
15	99	10.6	1593	16	08FY73
16					08FY73 pseudomonas

17	98.5	10.5	832	2	054356	054356 moraxella c
18	98	10.5	595	2	09S0U1	09S0U1 escherichia
19	98	10.5	595	16	08FCM2	08FCM2 escherichia
20	97.5	10.4	729	2	09KH34	09KH34 antarctic b
21	97.5	10.4	3165	16	08XY13	08XY13 raietonia s
22	97.5	10.4	4106	16	08XOP2	08XOP2 raietonia s
23	97	10.4	443	6	028243	028243 canis famill
24	96	10.3	356	12	08V0L7	08V0L7 equine herp
25	96	10.3	389	12	08V0M0	08V0M0 equine herp
26	96	10.3	826	12	08V0L5	08V0L5 equine herp
27	96	10.3	4776	16	097P71	097P71 streptococ
28	95.5	10.2	744	3	08TFG9	08TFG9 schizosacch
29	95.5	10.2	881	2	09AF09	09AF09 cellivibrio
30	95	10.1	122	13	08AY76	08AY76 meileagris g
31	94.5	10.1	438	16	08G3W5	08G3W5 bifidobacte
32	94.5	10.1	507	16	050506	050506 streptomyc
33	94.5	10.1	834	16	08DRJ3	08DRJ3 streptococ
34	94.5	10.1	3381	2	09KXJ3	09KXJ3 streptococ
35	94	10.0	570	2	059665	059665 pseudomonas
36	94	10.0	570	2	09AL29	09AL29 escherichia
37	94	10.0	967	3	008294	008294 saccharomyc
38	94	10.0	2190	16	08PGS0	08PGS0 xanthomonas
39	93.5	10.0	642	16	08XSW6	08XSW6 raietonia s
40	93.5	10.0	729	2	09EY54	09EY54 vibrio sp.
41	93	9.9	1236	3	09C105	09C105 schizosacch
42	92.5	9.9	1018	16	09HNU6	09HNU6 pseudomonas
43	92.5	9.9	1113	16	08FXA7	08FXA7 brucella su
44	92.5	9.9	1461	16	08XD19	08XD19 escherichia
45	92	9.8	316	12	08V0M4	08V0M4 equine herp

ALIGNMENTS

RESULT 1

08KXW2 PRELIMINARY; PRT; 179 AA.

AC 08KXW2; 01-OCT-2002 (TREMBLrel. 22, Created)

DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)

DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

DE Hypothetical protein.

OS Rhizobium etli.

OC Plasmid symbiotic plasmid p42d.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Rhizobiaceae; Rhizobium.

OX NCBI_TaxID=29449;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CFN42;

RX MEDLINE=91193195; PubMed=2013564;

RA Girard M.L., Flores M., Brom S., Romero D., Palacios R., Davila G.;

RT "Structural complexity of the symbiotic plasmid of Rhizobium

RT leguminosarum bv. phaseoli.";

RT J. Bacteriol. 173:2411-2419(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CFN42;

RX MEDLINE=97419521; PubMed=9274036;

RA Ramirez-Romero M.A., Bustos P., Girard L., Rodriguez O.,

RA Cevallos M.A., Davila G.;

RT "Sequence, localization and characteristics of the replicator region

RT of the symbiotic plasmid of Rhizobium etli.";

RT Microbiology 143:2825-2831(1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=CFN42;

RA Quintero V., Cevallos M.A., Davila G.;

RT "A site-specific recombinase and RecA are required to exert

RT incompatibility towards the symbiotic plasmid of Rhizobium etli.";

RT Submitted (JUN-2002) to the EMBL/Genbank/DBJ databases.

DR EMBL; U09028; AAM5018.1; -.

KW Hypothetical protein; Plasmid.
SQ SEQUENCE 179 AA; 18919 MW; 89F3E2783DCA266E CRC64;

Query Match 18.0%; Score 168.5; DB 2; Length 179;
Best Local Similarity 31.4%; Pred. No. 9e-05;
Matches 50; Conservative 22; Mismatches 70; Indels 17; Gaps 5;

QY 3 KIYVAIAIVSLTVSITASADPSKDSKAQVSAEAGITGTWYNQSGFTVTAGADGA 62
DB 23 KTMIRTNAAVALFAVATSLAFDASNFDPSSIASA--SSWQNSGSGTMIIOVDSFGN 80
QY 63 LRTGT---ESAVNAESRYLTGRYDSPAATDSSGTLAMQYAMKNNYNAHSATTWG- 118
DB 81 VSGQYNNRAGCTGCONSPYPLGRV-----NGTFIFSVGMNNSTENCNSATGWTGY 132
QY 119 QYVGAEARINTQWLTSGTTEANAWKSTLVGHDTFTKV 157
DB 133 AQVNGNNTETVISMNL--AYEGSGPALIEQGDFTQYV 168

RESULT 2

ID Q98SH4 PRELIMINARY; PRT; 152 AA.
AC Q98SH4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Avidin precursor.
GN AVD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Ahlroth M.K.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21070478; PubMed=11167523;
RA Ahlroth M.K., Kola E.H., Ewald D., Masabanda J., Sazanov A., Fries R.,
RA Kulomaa M.S.;
RT "Characterization and Chromosomal Localization of the Chicken Avidin
RT Gene Family.";
RT Annu. Genet. 31:367-375(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95394357; PubMed=7665080;
RA Wallen M.U., Laakkonen M.O., Kulomaa M.S.;
RT "Cloning and sequencing of the chicken egg-white avidin-encoding gene
RT and its relationship with the avidin-related genes Avrl-Avr5.";
RL Gene 161:205-209(1995).
DR EMBL, AJ311647; CAC34569.1;
DR HSSP; P02701; IAVD.
DR InterPro; IPR005469; Avidin.
DR InterPro; IPR005468; Avidin/str.
DR Pfam; PF01382; Avidin; 1.
DR PRINTS; PRO0709; AVIDIN.
DR PROSITE; PS00577; AVIDIN, 1.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 152 AA; 16742 MW; 04B4965B670B974B CRC64;

Query Match 15.5%; Score 145.5; DB 13; Length 152;
Best Local Similarity 33.3%; Pred. No. 0.0032;
Matches 44; Conservative 19; Mismatches 56; Indels 13; Gaps 7;

QY 34 VSAAGITGTWYNQSGFTFVTA-GADGALTGYESAVNAESRYLTGRYDSPAATDG 92
DB 23 LSARCSLTGKWDNDLGSNMTIGAVNSKGEFTGYTAV-TATSNIEIKESPLHGTONIN 81
QY 93 SGT--ALGWTYAMKNNYNAHSATTWGQYV---GGAERINTQWLTSGTTE-ANAWKS 146

DB 82 KRTQPTFGFTVNWK-----FSESTTVFGQCFIDRNKEV-LKTWMLLRSSVNDIGDQWKA 136
QY 147 TLVGHDTFTKVX 158
DB 137 TRVGINTFTRLR 148

RESULT 3

ID Q25058 PRELIMINARY; PRT; 529 AA.
AC Q25058;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel..23, Last annotation update)
DE Fibropellin Ia (Fragment).
OS Helicidaris erythrogramma (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinometridae;
OC Helicidaris.
OX NCBI_TaxID=7634;
RN [1]
RP SEQUENCE FROM N.A.
RA Bisgrove B.W.;
RT "Evolution of the fibropellin gene family and pattern of fibropellin
RT gene expression in sea urchin phylogeny.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; L33861; AAA29995.1;
DR HSSP; P08709; 18P9.
DR InterPro; IPR000152; Axx hydroxyl.
DR InterPro; IPR005469; Avidin.
DR InterPro; IPR005468; Avidin/str.
DR InterPro; IPR00742; EGF 2.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR001438; EGF II.
DR InterPro; IPR006209; EGF-like.
DR Pfam; PF01382; Avidin; 1.
DR Pfam; PF00008; EGF; 10.
DR PRINTS; PR00709; AVIDIN.
DR PRINTS; PR00010; EGFBLDOD.
DR SMART; SM00179; EGF CA; 9.
DR PROSITE; PS00010; Axx hydroxyl; 8.
DR PROSITE; PS00577; AVIDIN; 1.
DR PROSITE; PS00022; EGF 1; 11.
DR PROSITE; PS01186; EGF 2; 10.
DR PROSITE; PS01187; EGF CA; 7.
KW EGF-like domain.
FT NON-TER.
SQ SEQUENCE 529 AA; 55543 MW; D4AB958FCF9ACB5A CRC64;

Query Match 12.2%; Score 114; DB 5; Length 529;
Best Local Similarity 28.4%; Pred. No. 2.4;
Matches 40; Conservative 22; Mismatches 55; Indels 24; Gaps 7;

QY 34 VSAEAG---ITGTWYNQSGFTVTAAGDALTGY-----ESAVNAESRYLT---TGR 83
DB 396 VNCDEVGFCDLEGWVYECNDQITITKSTGMILGYMTAVEAVGYAAPTVAAGVGAANN 455
QY 84 YDSAPATDGSGLTGWTVAMKNNYRNAHSATTWGQ-YVGAEARINTQWLTSGT-TEA 141
DB 456 YDF-----PSFGFTVAVDNGM-----STTSMTAQCHLCDNEEVLTYTWINNMVDTQ 503
QY 142 NAWKSTLVGHDTFTKVPSAA 162
DB 504 DIKSNMVGQDKMTRYEQSIA 524

RESULT 4
ID Q06353 PRELIMINARY; PRT; 595 AA.
AC Q06353;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)


```

RESULT 7
Q9S0T7 ID Q9S0T7 PRELIMINARY; PRT; 595 AA.
AC Q9S0T7;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Flagellin.
GN Flagellin.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bi 316-42;
RA Ohnishi K., Ishioka K., Matsuba T., Harayama S.;
RT "Cloning of H antigen genes in E.coli serotypes and expression in
RT E.coli K-12.",
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB028475; BAA85084.1; -
DR InterPro; IPR001492; Flagellin_N.
DR InterPro; IPR001029; Flagellin_C.
DR Pfam; PF00700; Flagellin_N; 1.
DR PRINTS; PR00207; FLAGELLIN.
DR PRODOM; PD000316; Flagellin_C; 2.
SQ SEQUENCE 595 AA; 61020 MW; 8B9DC3D6ACT8C427 CRC64;

Query Match 11.1%; Score 104; DB 2; Length 595;
Best Local Similarity 25.6%; Pred. No. 14;
Matches 51; Conservative 30; Mismatches 88; Indels 30; Gaps 9;

QY 7 AAIASVLTITVSITASADPSKSKAQAASAEAGITGTWYNOLGST---FIYTAGADGAL 63
DB 343 ASVTMGCTTYNFTGTGADAD-AAATNANGVSFTDTASKETLVNKAVATKCKGKAVAAADGDTSA 401
QY 64 TGTYESAV-----GMAESRYVLTGRYDSAPAT---DSSGTAIG-WTVAMK--- 104
DB 402 TTTYSKGVOTYQAVFAPAGDGTASAKYADKADVDVSNATATYTDADGEMTTTIGSYTTKYSIDA 461
QY 105 NNYRNASHATTWSGOVVG--GAEARINTQWLTS-GTTEANMKSTLVHD---TFTKV 157
DB 462 NNGKTVDSGTGTGAKAPKVGAVVYANGTLTTATSGVTVKDPLKALDAISSIDKF 521
QY 158 KPSAASIDAKKAKGAVNGN 176
DB 522 RSSLGIHQRLDSAVTNLN 540

RESULT 8
Q8Y4J4 ID Q8Y4J4 PRELIMINARY; PRT; 1310 AA.
AC Q8Y4J4;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein lmo2444.
GN LMO2444.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rusnlok C., Amend A.,
RA Bagneri F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Dommann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Enttlan K.-D., Fsthi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Geobel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz U., Kuhn M., Kunert F., Kutzkat G.,

```

```

RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordisiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluerer T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species.",
RL Science 294:849-852(2001).
DR EMBL; AL591983; CAD005522.1; -
DR Listlist; LMO2444; -
DR InterPro; IPR005084; CBM 6.
DR InterPro; IPR00322; Glyco_hydro_31.
DR Pfam; PF03422; CBM_6; 1.
DR Pfam; PF01055; Glyco_hydro_31; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1310 AA; 144058 MW; D94919CA04021583 CRC64;

Query Match 10.9%; Score 102; DB 16; Length 1310;
Best Local Similarity 24.9%; Pred. No. 50;
Matches 48; Conservative 23; Mismatches 52; Indels 70; Gaps 9;

QY 12 SLTTSITASADPSKSKAQAASAEAGITGTWYNOLGSTFIYTAGADGAL-TGTYESA 70
DB 860 AVTYVKVPAAQADAD-----TVITSSASIVOTGTYEME 893
QY 71 VGN-----AESRYVLTGRYDSAPATDGSCTALGWTVMKN-----YRNAS- 112
DB 894 TGSNFPADTVADKPVAEKTYV--DGYDK-----DGAGTITVANVKDSDGVNVDLTYSKASD 947
QY 113 ---ATTWSGOVGAEARINTQWLTS-GTTEANMKSTL---VGHDTFTKV----- 157
DB 948 NQALSTFYNGRYVQOTTLKPTWDSVSETPLSAGNKSISYKVTDTGKADQVSLDKV 1007
QY 158 ---KPSAASIDA 166
DB 1008 NIGFTPTVAKVEA 1020

RESULT 9
Q8XQ42 ID Q8XQ42 PRELIMINARY; PRT; 1371 AA.
AC Q8XQ42;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Putative hemagglutinin-related protein.
GN RSP1444 OR R803099.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plasmid megaplasmid.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brotilier P., Camus J.C., Catalioco L.,
RA Chandlier M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaepin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.",
RL Nature 415:497-502(2002).
DR EMBL; AL646084; CAD18595.1; -
DR InterPro; IPR006172; DNA_pol_B.
DR PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW Plasmid; Complete proteome.
SQ SEQUENCE 1371 AA; 132558 MW; 1D94A3037BF8FC8F CRC64;

Query Match 10.9%; Score 102; DB 16; Length 1371;
Best Local Similarity 23.7%; Pred. No. 53;
Matches 52; Conservative 26; Mismatches 61; Indels 80; Gaps 9;

QY 8 AAIASVLTITVSITASADPSKSKAQAASAEAGITGTWYNOLGSTFIYTAGADGAL-TGTYESA 70

```

```

Db 685 AILLTGATGATATMTWAGSVTLALDLSLTSAGVATISGTSANVSSRYRQGLSTVINAGT 744
QY 58 GADGA-----LTGTYESAVGNAESRYVLTGRYDSAPATDG-----SGTA 96
Db 745 GSNAAASLTGCTVITGT-NSVGNASS-----TSNGNAAVLTDGKVLTTATACPINIAGSN 799
QY 97 LGMTVAMKN-----NYRNAHSAITWMSGGVGGAERINTQWLTSCTT 139
Db 800 AGDGGVWSSGSAVMTSAPASSSITTSARSLDSVSGISGFYIGGSGKTL----- 848
QY 140 EANAAMKSTLVGHDFTFKVP---SASIDAKKAGVNG 175
Db 849 -----TFATTAARVSTIAESLVGARAKAFNMKG 874

RESULT 10
Q8YJ4M4 PRELIMINARY; PRT; 1374 AA.
ID O8YJ4M4;
AC O8YJ4M4;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Extracellular serine protease (EC 3.4.21.-).
GN BMEI0058.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OC NCB1_TaxID=29459;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lyfidis A., Reznik G.,
RA Jabloncki L., Larsen N., D'Souza M., Bernal A., Mazur M., Goleman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Lelsson J.-J.,
RA Haselkorn R., Kyriides N., Overbeek R.,
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL EMBL; AEO09449; AAL51240.1; ".
DR InterPro: IPR006315; Autotransporter.
DR InterPro: IPR005546; Autotransporter.
DR Pfam: PF03797; Autotransporter; 1.
DR TIGRFAMs: TIGR01414; autotrans_bar1. 2.
DR HydroLase: Complete proteome.
SQ SEQUENCE 1374 AA; 140577 MW; AE65C4BD5F60AB3 CRC64;

Query Match 10.8%; Score 102; DB 16; Length 1374;
Best Local Similarity 21.5%; Pred. No. 53;
Matches 45; Conservative 29; Mismatches 65; Indels 70; Gaps 7;

QY 20 ASASADPSKDSKAQVSAEAGI---TGTWYNOLGS-----TFIVTAGAD---- 60
Db 1078 APAGADPASSPVAGVYTSNDGIMARIGDYSKLOSSLTWNQNTRTVLIQSGVGKPY 1137
QY 61 ----GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGW----- 99
Db 1138 EADTGKILGGINLALYGAISRI-----NSPFGDGAITTSAMGLGTLTWGESGFYVD 1190
QY 100 ----TWAMKNYRNAHS-----ATTWSGVGVGAERINTQWLTT-----S 136
Db 1191 GAAQINWYNNDYSDTGTGKLADDKATGYAVSIETGQRNIGERWSVTPOAQLMWSKLS 1250
QY 137 GTTEANAMKSTLVGHDFTFKVKSASID 165
Db 1251 MDTFNNIWEANVSLNDSLSLRAGVALD 1279

RESULT 11
Q06351 PRELIMINARY; PRT; 595 AA.
ID Q06351

```

```

AC 006351;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Flagellin.
GN FlIC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCB1_TaxID=562;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Su 1242 (Orskov);
RA Schoenhalz G.J.;
RL Thesis (1992), Unknown Institution.
RN [2]_
RP SEQUENCE FROM N.A.
RC STRAIN=Su 1242 (Orskov);
RX MEDLINE=93374833; PubMed=8366026;
RA Schoenhalz G.J., Whitfield C.,
RT "Comparative analysis of flagellin sequences from Escherichia coli
RT strains possessing serologically distinct flagellar filaments with a
RT shared complex surface pattern."
RL J. Bacteriol. 175:5395-5402(1993).
DR EMBL; L07387; AAA23797.1; ".
DR InterPro: IPR001492; Flagellin.
DR InterPro: IPR001029; Flagellin_C.
DR Pfam: PF00700; Flagellin_C; 1.
DR Pfam: PF00669; Flagellin_N; 1.
DR PRINTS: PR00207; FLAGELLIN.
DR ProDom: PD000316; Flagellin_C; 2.
SQ SEQUENCE 595 AA; 60923 MW; 9AAFB2E86884607A CRC64;

Query Match 10.8%; Score 101; DB 2; Length 595;
Best Local Similarity 25.6%; Pred. No. 24;
Matches 51; Conservative 29; Mismatches 89; Indels 30; Gaps 9;

QY 7 AAIANSLTIVSTTASASADPSKDSKAQVSAEAGITGTYNOLGSR---FIYTAGDGL 63
Db 343 ASVTMGCTTYNFKTGADAG-AATNAGVSFTDTASKETVLNVAATKAGTAAANGDTSA 401
QY 64 TGTYESAV-----GNAESRYVLTGRYDSAPAT-----DGSCTALG-WTVAMK--- 104
Db 402 TITTSAGVTVQAVPAAGGTASAKYADNTVSNATATYTDADGEMTTIGSYTTKXSIDA 461
QY 105 NNRYNAHSAITWMSGGVG--GAERINTQWLTS-GTTEANAMKSTLVGHD---FTTKV 157
Db 462 NNGKVTVDGSGKYAPVGAEVVVSANGTLTDTATSEGTVKDPLKALDAISSIDKF 521

QY 158 KPSASIDAKKAGVNGN 176
Db 522 RSLGAIQNRLDLSAVTNLN 540

RESULT 12
Q9S4M2 PRELIMINARY; PRT; 560 AA.
ID Q9S4M2;
AC Q9S4M2;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Flagellin (Fragment).
GN FlIC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OC NCB1_TaxID=562;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=E74/68;
RX MEDLINE=9084952; PubMed=9864325;
RA Reid S.D., Selander R.K., Whitam T.S.;
RT "Sequence diversity of flagellin (flc) alleles in pathogenic

```

RT Escherichia coli.;
 RL U. Bacteriol. 181:153-160(1999).
 DR EMBL; AF128954; AAD28525.2; -;
 DR InterPro; IPR001492; Flagellin_N.
 DR InterPro; IPR001029; Flagellin_C.
 DR Pfam; PF00700; Flagellin_C; 1.
 DR Pfam; PF00669; Flagellin_N; 1.
 DR PRINTS; PR00207; FLAGELLIN.
 DR Prodom; PD000316; Flagellin_C; 2.
 FT NON_TER 1
 FT NON_TER 560
 SQ SEQUENCE 560 AA; 57262 MW; 97DF005C9E0AB71 CRC64;

Query Match 10.7%; Score 100; DB 2; Length 560;
 Best Local Similarity 25.1%; Pred. No. 26;
 Matches 50; Conservative 30; Mismatches 89; Indels 30; Gaps 9;

QY 7 AAIIVSLTVSITASASADPSKQVSAEAGITGTWYNOLGST---FIVTAGADAL 63
 DB 324 ASVTMGITVYKRTGADAD-AATNAGVFTPTASKETVLNKVATKAGKAAADGDTSA 382
 QY 64 TGTYESAV-----GNAESRYVLTGRYDSAPAT---DGSGTALG-WTVAMK--- 104
 DB 383 TITVYSGVQYQAVFAAGDGTASAKADADVSNATATYTDADGEMTTIGSYTTKYSIDA 442
 QY 105 NNRYRHAHTTWSGQYV--GAERINTQWLTS-GTTEANAMKSTLVGHD---TFPKV 157
 DB 443 NNGKVTVDSGTGTGKAPKVAEYVYVANGTLTDTATSEGTVPKPLKALDEAISISDKF 502
 QY 158 KPSASIDAKKAGVNNNGN 176
 DB 503 RSSLGAIQNRDLSAVTNLN 521

RESULT 13
 Q9S4M4 PRELIMINARY; PRT; 565 AA.

AC Q9S4M4; 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Flagellin (Fragment).
 GN FLIC.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=DEC 6a;
 RX MEDLINE=99084952; PubMed=9864325;
 RA Reid S.D., Selander R.K., Whitlam T.S.;
 RT "Sequence diversity of flagellin (flc) alleles in pathogenic
 RT Escherichia coli."
 RL J. Bacteriol. 181:153-160(1999).
 DR EMBL; AF128952; AAD28525.2; -;
 DR InterPro; IPR001492; Flagellin_N.
 DR InterPro; IPR001029; Flagellin_C.
 DR Pfam; PF00700; Flagellin_C; 1.
 DR Pfam; PF00669; Flagellin_N; 1.
 DR PRINTS; PR00207; FLAGELLIN.
 DR Prodom; PD000316; Flagellin_C; 2.
 FT NON_TER 1
 FT NON_TER 565
 SQ SEQUENCE 565 AA; 57857 MW; 6E9E8A1AF72DC29 CRC64;

Query Match 10.7%; Score 100; DB 2; Length 565;
 Best Local Similarity 25.1%; Pred. No. 26;
 Matches 50; Conservative 30; Mismatches 89; Indels 30; Gaps 9;
 QY 7 AAIIVSLTVSITASASADPSKQVSAEAGITGTWYNOLGST---FIVTAGADAL 63
 DB 324 ASVTMGITVYKRTGADAD-AATNAGVFTPTASKETVLNKVATKAGKAAADGDTSA 386

QY 64 TGTYESAV-----GNAESRYVLTGRYDSAPAT---DGSGTALG-WTVAMK--- 104
 DB 387 TITVYSGVQYQAVFAAGDGTASAKADADVSNATATYTDADGEMTTIGSYTTKYSIDA 446
 QY 105 NNRYRHAHTTWSGQYV--GAERINTQWLTS-GTTEANAMKSTLVGHD---TFPKV 157
 DB 447 NNGKVTVDSGTGTGKAPKVAEYVYVANGTLTDTATSEGTVPKPLKALDEAISISDKF 506
 QY 158 KPSASIDAKKAGVNNNGN 176
 DB 507 RSSLGAIQNRDLSAVTNLN 525

RESULT 14
 Q48028 PRELIMINARY; PRT; 1477 AA.

AC Q48028; 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Adhesin.
 GN HMM2A.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=12;
 RX MEDLINE=92192797; PubMed=1548058;
 RA Barenkamp S.U., Leininger E.;
 RT "Cloning, expression, and DNA sequence analysis of genes encoding
 RT nontypeable Haemophilus influenzae high-molecular-weight surface-
 RT exposed proteins related to filamentous hemagglutinin of Bordetella
 RT pertussis."
 RL Infect. Immun. 60:1302-1313(1992).
 DR EMBL; U08875; AAA20524.1; -;
 DR InterPro; IPR001000; Glyco_hydro.10.
 DR InterPro; IPR000169; SHprot acate.
 DR PROSITE; PS00591; GLYCOSYL_HYDROL_F10; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
 SQ SEQUENCE 1477 AA; 154473 MW; B057C23F1AD24B0E CRC64;

Query Match 10.6%; Score 99.5; DB 2; Length 1477;
 Best Local Similarity 26.2%; Pred. No. 87;
 Matches 50; Conservative 26; Mismatches 88; Indels 27; Gaps 6;

QY 11 VSLTIVSITASASADPSKQVSAEAGI---GTWYNOL-GSTFIVTAGADALTGT 66
 DB 1177 ISGNTVSASAGDLTTRKSGKIEKSGEANTVATGTIGTISGNTVNVANA-GDLTVG 1235
 QY 67 YESAVNGAESRYVLTGRYDSAPATDGS-----TALGVAMKNNRYRHAHTTW 116
 DB 1236 NGAEINATGGAATLTATGNTITTTAGSSITSTKQVULLAONGSIAGSINAAVTLNITG 1235
 QY 117 SGQYVGAEARINTQWLTS-GTTEANAMKSTLVGHDFTFKVPSA-----ASIDAAKA 170
 DB 1296 TLTIVASGDIK-----ATSGTLVINAKDAKNDAGDSTEYVAVNAVSSGSVTAATSS 1349
 QY 171 GYNNNGNPLDAY 181
 DB 1350 SVNITGDLNTV 1360

RESULT 15

Q912M3 PRELIMINARY; PRT; 2468 AA.
 ID Q912M3; 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Hypothetical protein PA1874.

GN PA1874.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OK NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 Brody L.L., Coulter S.N., Folger K.R., Kae A., Larbig K., Lim R.M.,
 Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL; AE004613; AG05263.1; .
 DR InterPro: IPR001343; Hemlysn_Ca_bind.
 DR InterPro: IPR006162; Ppanthe_atlch.
 DR PRINTS; PRO0313; CABNNGRPT.
 DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 2468 AA; 238414 MW; 13596AFAB2C4B899 CRC64;

Query Match 10.6%; Score 99.5; DB 16; Length 2468;
 Best Local Similarity 24.9%; Pred. NO.1.6e+02;

Matches 61; Conservative 19; Mismatches 70; Indels 95; Gaps 12;

QY 14 TTVSITASASA-----DPSKDSKAQVSAEAG-ITGTWYNQLGSTPIVAG--- 58
 DB 1381 TVNVNTASDAAGTSLPATTVDSSLPSIPQVDPSPNGSVISGT--ADAGTIIITDNGN 1438
 QY 59 -----ADGA-----LTGTYESAVGNAESRYVLTGRYDSAPA----- 89
 DB 1439 PIGQVTVADSGSNWSTPGIPLPDGTVVNVVARSPS-----NVDSAPAVITVDGVAPAAP 1492
 QY 90 -----TDGSGTA-LGWTYAMKNNRNNAHSATTWGGVYGSEARINTQWLLTSGTEA 141
 DB 1493 VIDPSNGTEISGTAEAGATVI-----LTDGGGNPIGQATADSGNWTFTPTSTPLA 1542
 QY 142 NAWKSTLVGHDTFTKYK-PSAASIDA-----AKKAGV-----NNGN 176
 DB 1543 NCTVINAVADPAGNVTGPRASVTVDALAPPAPVINSNGVVISGTAAGATVILTDNGN 1602
 QY 177 PLDAV 181
 DB 1603 PIGQV 1607

Search completed: October 27, 2003, 10:43:17
 Job time : 43 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993.- 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:40:04 ; Search time 11 Seconds

(without alignments)
782.353 Million cell updates/sec

Title: US-09-589-870B-2

Sequence: 1 MKRIVVAIAVSLTIVSITR.....IDAKKAGVNGNPLDVAQO 183

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	936	100.0	183	1 SAV_STRV	P22629 streptomyc
2	932	99.6	183	1 SAVI_STRVL	O53532 streptomyc
3	897	95.8	183	1 SAV2_STRVL	O53533 streptomyc
4	146.5	15.7	150	1 AVR4_CHICK	P56734 gallus gall
5	143.5	15.3	152	1 AVID_CHICK	P02701 gallus gall
6	124.5	13.3	150	1 AVR6_CHICK	P56735 gallus gall
7	121.5	13.0	150	1 AVR1_CHICK	O13153 gallus gall
8	120.5	12.9	150	1 AVR7_CHICK	P56736 gallus gall
9	120.5	12.9	150	1 AVR3_CHICK	P56737 gallus gall
10	120	12.8	164	1 FBPI_STRPU	P10079 stronglyl
11	116.5	12.4	150	1 AVR2_CHICK	P56732 gallus gall
12	100.5	10.7	1025	1 SLAP_CAVCR	P35828 caulobacter
13	96.5	10.3	741	1 MAS2_MYCTU	O50596 mycobacter
14	93	9.9	1225	1 Y309_MYCTU	P47551 mycoplasma
15	91.5	9.8	1120	1 STRP_ECOLI	P76072 escherichia
16	91.5	9.8	2249	1 OMPA_RICRI	P15921 rickettsia
17	90	9.6	1150	1 APNU_PIG	P12021 sus scrofa
18	89	9.5	255	1 CBPM_STRAL	P00773 streptomyc
19	89	9.5	303	1 HFAD_CAVCR	O45977 caulobacter
20	89	9.5	488	1 PHB_ALCPA	P12625 alcaligenes
21	88.5	9.5	455	1 SLAP_LACBR	O05044 lacobacilli
22	87.5	9.3	457	1 PRTC_STRGR	P52320 streptomyc
23	87	9.3	666	1 MUR2_ENTHR	P39046 enterococci
24	86.5	9.2	500	1 ABR1_TRIRE	O92455 trichoderma
25	86	9.2	598	1 VAUF_SCHPO	O10168 schizosacch
26	85	9.1	797	1 VGLX_HSVB	P28968 equine herp
27	85	9.1	872	1 GUXA_CELFI	P50401 cellulomona
28	84.5	9.0	962	1 GUNA_PSEFL	P10426 pseudomonas
29	83.5	8.9	500	1 XYLA_TRIKO	P48792 trichoderma
30	83.5	8.9	575	1 FLAB_CAMCE	P22251 campylobact
31	83.5	8.9	575	1 FLAB_CAMCE	P22251 campylobact
32	83.5	8.9	917	1 ICAS_MOUSE	O60655 mus musculu
33	83.5	8.9	1645	1 OMPB_RICTY	P96989 r outer mem

34	83.5	8.9	2021	1 OMPA_RICCN	O52657 rickettsia
35	83	8.9	610	1 CHIT_STRPL	P11220 streptomyc
36	82.5	8.8	297	1 PRTA_STRGR	P00776 streptomyc
37	82.5	8.8	331	1 ADT1_WHEAT	O41629 triticum ae
38	82.5	8.8	563	1 HEMA_IACKV	P09343 influenza a
39	82.5	8.8	616	1 MUTA_STRCM	O05064 streptomyc
40	82	8.8	381	1 ADT1_ARATH	P31167 arabidopsis
41	82	8.8	511	1 GUNA_PSEFL	P18126 pseudomonas
42	82	8.8	570	1 FBPI_STRPU	P49013 stronglyl
43	82	8.8	572	1 FLAB_CAMCO	P27053 campylobact
44	82	8.8	572	1 FLAB_CAMCO	P18245 campylobact
45	82	8.8	704	1 OEB6_NPVAC	O00704 autocographa

ALIGNMENTS

RESULT 1
ID SAV_STRV STANDARD: PRT; 183 AA.
AC P22629;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Streptavidin precursor.
OS Streptomyces avidinii.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1895;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-64.
RX MEDLINE=86148514; PubMed=3951999;
RA Argarana C.E., Kuntz I.D., Birken S., Axel R., Cantor C.R.;
RL "Molecular cloning and nucleotide sequence of the streptavidin gene."; Nucleic Acids Res. 14:1871-1882(1986).
RN [2]
RP IMPORTANCE OF TRP IN BIOTIN-BINDING.
RX MEDLINE=89134083; PubMed=3223904;
RA Gitlin G., Bayer E.A., Wilchek M.;
RL "Studies on the biotin-binding site of streptavidin. Tryptophan residues involved in the active site."; Biochem. J. 256:279-282(1988).
RN [3]
RP IMPORTANCE OF TYR IN BIOTIN-BINDING.
RX MEDLINE=90351377; PubMed=2386489;
RA Gitlin G., Bayer E.A., Wilchek M.;
RL "Studies on the biotin-binding sites of avidin and streptavidin. Tyrosine residues are involved in the binding site."; Biochem. J. 269:527-530(1990).
RN [4]
RP CELL-BINDING.
RX MEDLINE=90358825; PubMed=2390089;
RA Alon R., Bayer E.A., Wilchek M.;
RL "Streptavidin contains an RVD sequence which mimics the RGD receptor domain of fibronectin."; Biochem. Biophys. Res. Commun. 170:1236-1241(1990).
RN [5]
RP X-RAY CRYSTALLOGRAPHY OF 37-157.
RX MEDLINE=97337436; PubMed=9194176;
RA Weber P.C., Orlowski D.H., Wendoloski J.J., Selenka F.R.;
RL "Structural origins of high-affinity biotin binding to streptavidin."; Science 243:85-88(1989).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 40-157.
RX MEDLINE=97337436; PubMed=9194176;
RA Freitag S., le Trong I., Klumb L., Steylen P.S., Stenkamp R.E.;
RL "Structural studies of the streptavidin binding loop."; Protein Sci. 6:1157-1166(1997).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.45 ANGSTROMS) OF 37-157.
RX MEDLINE=97294734; PubMed=9148939;
RA Katz B.A., Cases R.T.;
RL "In crystals of complexes of streptavidin with peptide ligands

RT containing the HPQ sequence the pKa of the peptide histidine is less
than 3.0.";
J. Biol. Chem. 272:13220-13228(1997).
[8]
RN X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 37-157.
RP MEDLINE=98070615; PubMed=9405158;
RA Katz B.A.;
RT "Binding of biotin to streptavidin stabilizes intersubunit salt
bridges between Asp61 and His87 at low pH.";
J. Mol. Biol. 274:776-800(1997).
[9]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 37-157 OF MUTANTS.
RA MEDLINE=98300340; PubMed=9636711;
RA Freltag S., Le Trong I., Chikoti A., Klumb L.A., Stayton P.S.,
Stenkamp R.E.;
RT "Structural studies of binding site tyrophophan mutants in the high-
affinity streptavidin-biotin complex.";
J. Mol. Biol. 279:211-221(1998).
CC - FUNCTION: THE BIOLOGICAL FUNCTION OF STREPTAVIDIN IS NOT KNOWN.
CC - FORMS A STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE
MOLECULE OF BIOTIN PER SUBUNIT OF STREPTAVIDIN).
CC - SUBUNIT: Homotetramer.
CC - SUBCELLULAR LOCATION: Secreted.
CC - SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC - DATABASE: NAME=ProZyme technical fact sheet;
WWW="http://www.prozyme.com/technical/sa10data.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X03591; CAA37265.1; -
DR PIR, A23513; A23513.
DR PDB, 1STP, 15-OCT-94.
DR PDB, 1PTS, 15-OCT-94.
DR PDB, 1SRE, 01-NOV-94.
DR PDB, 1SRF, 01-NOV-94.
DR PDB, 1SRG, 01-NOV-94.
DR PDB, 1SRH, 01-NOV-94.
DR PDB, 1SRI, 01-NOV-94.
DR PDB, 1SRJ, 01-NOV-94.
DR PDB, 1RST, 15-MAY-97.
DR PDB, 1RSU, 15-MAY-97.
DR PDB, 1STD, 03-APR-96.
DR PDB, 1SLE, 03-APR-96.
DR PDB, 1SLF, 03-APR-96.
DR PDB, 1SLG, 03-APR-96.
DR PDB, 1STR, 08-MAR-96.
DR PDB, 1STS, 08-MAR-96.
DR PDB, 1SWA, 04-MAR-98.
DR PDB, 1SWB, 04-MAR-98.
DR PDB, 1SWC, 04-MAR-98.
DR PDB, 1SWD, 04-MAR-98.
DR PDB, 1SWE, 04-MAR-98.
DR PDB, 1SWG, 15-JUL-98.
DR PDB, 1SWH, 23-MAR-99.
DR PDB, 1SWI, 23-MAR-99.
DR PDB, 1SWJ, 23-MAR-99.
DR PDB, 1SWK, 23-MAR-99.
DR PDB, 1SWL, 23-MAR-99.
DR PDB, 1SWM, 23-MAR-99.
DR PDB, 1SWN, 23-MAR-99.
DR PDB, 1SWO, 23-MAR-99.
DR PDB, 1SWP, 23-MAR-99.
DR PDB, 1SWQ, 23-MAR-99.
DR PDB, 1SWR, 23-MAR-99.
DR PDB, 1SWA, 18-MAR-98.
DR PDB, 1VWA, 18-MAR-98.
DR PDB, 1VWC, 18-MAR-98.
DR PDB, 1VWD, 18-MAR-98.
DR PDB, 1VWE, 18-MAR-98.

DR PDB, 1VWF, 18-MAR-98.
DR PDB, 1VWG, 18-MAR-98.
DR PDB, 1VWH, 18-MAR-98.
DR PDB, 1VWI, 18-MAR-98.
DR PDB, 1VWJ, 18-MAR-98.
DR PDB, 1VWK, 18-MAR-98.
DR PDB, 1VWL, 18-MAR-98.
DR PDB, 1VWM, 18-MAR-98.
DR PDB, 1VWN, 18-MAR-98.
DR PDB, 1VWO, 18-MAR-98.
DR PDB, 1VWP, 18-MAR-98.
DR PDB, 1VWQ, 18-MAR-98.
DR PDB, 1VWR, 18-MAR-98.
DR PDB, 1SWF, 29-APR-98.
DR PDB, 21ZA, 16-SEP-98.
DR PDB, 21ZB, 16-SEP-98.
DR PDB, 21ZC, 16-SEP-98.
DR PDB, 21ZD, 23-SEP-98.
DR PDB, 21ZE, 16-SEP-98.
DR PDB, 21ZF, 16-SEP-98.
DR PDB, 21ZG, 16-SEP-98.
DR PDB, 21ZH, 16-SEP-98.
DR PDB, 21ZI, 16-SEP-98.
DR PDB, 21ZJ, 16-SEP-98.
DR PDB, 21ZK, 16-SEP-98.
DR PDB, 21ZL, 16-SEP-98.
DR PDB, 2RTA, 18-NOV-98.
DR PDB, 2RTB, 18-NOV-98.
DR PDB, 2RTC, 18-NOV-98.
DR PDB, 2RTD, 18-NOV-98.
DR PDB, 2RTE, 18-NOV-98.
DR PDB, 2RTF, 27-JAN-99.
DR PDB, 2RTG, 18-NOV-98.
DR PDB, 2RTH, 18-NOV-98.
DR PDB, 2RTI, 25-NOV-98.
DR PDB, 2RTJ, 25-NOV-98.
DR PDB, 2RTK, 18-NOV-98.
DR PDB, 2RTL, 18-NOV-98.
DR PDB, 2RTM, 18-NOV-98.
DR PDB, 2RTN, 18-NOV-98.
DR PDB, 2RTO, 18-NOV-98.
DR PDB, 2RTP, 18-NOV-98.
DR PDB, 2RTQ, 18-NOV-98.
DR PDB, 2RTR, 18-NOV-98.
DR PDB, 1DF8, 20-SEP-00.
DR PDB, 1I9H, 25-DEC-02.
DR PDB, 1KEF, 01-MAY-02.
DR PDB, 1KL3, 01-MAY-02.
DR PDB, 1KL4, 01-MAY-02.
DR PDB, 1KL5, 01-MAY-02.
DR PDB, 1LCV, 06-NOV-02.
DR PDB, 1LCW, 06-NOV-02.
DR PDB, 1LCZ, 06-NOV-02.
DR PDB, 1SMS, 30-JUL-99.
DR PDB, 1SWT, 15-DEC-99.
DR PDB, 1SWU, 17-NOV-99.
DR InterPro: IPR005469; Avidin.
DR InterPro: IPR005468; Avidin/str.
DR InterPro: IPR005470; Streptavidin.
DR Pfam, PF01382; Avidin, 1.
DR PRINTS, PR00709; Avidin.
DR PROSITE, PS00577; AVIDIN, 1.
KW Signal, Biotin; 3D-structure.
FT SIGNAL 1 24
FT CHAIN 25 183
FT BINDING 67 67
FT BINDING 78 78
FT BINDING 116 116
FT BINDING 132 132
FT BINDING 144 144
FT SITE 83 85
FT TURN 38 41
FT STRAND 43 47
STRAND
STREPTAVIDIN.
INVOLVED IN BIOTIN BINDING.
INVOLVED IN BIOTIN BINDING.
INVOLVED IN BIOTIN BINDING.
INVOLVED IN BIOTIN BINDING.
INVOLVED IN BIOTIN BINDING.
CELL ATTACHMENT SITE.


```

FT BINDING 78 78 SIMILARITY) INVOLVED IN BIOTIN BINDING (BY
FT CHAIN 1 INVOLVED IN BIOTIN BINDING (BY
FT BINDING 116 116 INVOLVED IN BIOTIN BINDING (BY
FT BINDING 132 132 INVOLVED IN BIOTIN BINDING (BY
FT BINDING 144 144 INVOLVED IN BIOTIN BINDING (BY
SQ SEQUENCE 183 AA; 18833 MW; FEAFFDFE4ECCA CRC64;

Query Match
Best Local Similarity 95.8%; Score 897; DB 1; Length 183;
Matches 174; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 MKRIVAAIAVSLTIVSITASADPSKSKQVSAEAGITGTWYNOLGTFITAGAD 60
DB 1 MKRIVAAIAVSLTIVGITAASADPSKSKQAAVAEAGITGTWYNOLGTFITANAD 60
OY 61 GALTGTYESAVGNASRYVLTGRYDAPATDGGTALGWTYAKNNYRNAHSATTWGGQY 120
DB 61 GSLTGTYESAVGNASRYVLTGRYDAPATDGGTALGWTYAKNNYRNAHSATTWGGQY 120
OY 121 VCGAARINTOWLLTSGTTEANAKSTLVGHDTFTKVKRSASIDAAKKAGVNNGNPDA 180
DB 121 VAGSEARINTOWLLTSGTTEANAKSTLVGHDTFTKVKRSASIDAAKKAGVNNGNPDA 180
OY 181 VQQ 183
DB 181 VQQ 183

RESULT 4
AVID CHICK STANDARD; PRT; 150 AA.
AC P56734;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Avidin-related protein 4/5 precursor.
GN AV4 AND AV5.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White Leghorn; TISSUE=Oviduct;
RX MEDLINE=94170814; PubMed=8125122;
RA Keinaenen R.A., Wallen M.J., Kristo P.A., Laukkanen M.O.,
RA Toimela T.A., Helenius M.A., Kulomaa M.S.;
RT "Molecular cloning and nucleotide sequence of chicken avidin-related
RT genes 1-5.";
RL Eur. J. Biochem. 220:615-621(1994).
CC -1- MISCELLANEOUS: THE SEQUENCES OF THE CODING REGIONS OF GENES AV4
CC AND AV5 ARE IDENTICAL.
CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z22883; -; NOT_ANNOTATED_CDS.
CC PIR; S42204; S42204.
CC HSSP; P02701; 1RAV.
CC InterPro; IPR005468; Avidin/str.
CC Pfam; PF01382; Avidin; 1.
CC PROSITE; PS00577; AVIDIN; FALSE_NEG.

```

```

KM Biotin; Signal; Multigene family.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 150 AVIDIN-RELATED PROTEIN 4/5.
FT BINDING 57 57 BIOTIN (NON-COVALENT) (BY SIMILARITY).
FT CARBOHYD 67 67 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 150 AA; 16644 MW; 9A6C6C6310EFEL3A CRC64;

Query Match
Best Local Similarity 15.7%; Score 146.5; DB 1; Length 150;
Matches 45; Conservative 22; Mismatches 47; Indels 19; Gaps 8;

OY 34 VSAEAGITGTWYNOLGTFITVA-GADGALTGTYESAV---GNAERYVLTGRYDSAP 88
DB 23 LSARKCSLTGKWTNNLGSIMTIRAVNSRGFTGYTLTAVADNPNTILSPGLGIQHKRA- 81
OY 89 ATDGGTALGWTYAKNNYRNAHSATTWGGQY---GGAERINTOWLLTSGTTEAN-AW 144
DB 82 ---SQPTFGFTVH--VF--SESTVFTGQCFIDRNGKEY-LKTMMLRSSVNDISYDW 132
OY 145 KSTLVGHDTFTKV 157
DB 133 KATRVGNVNFTRL 145

RESULT 5
AVID CHICK STANDARD; PRT; 152 AA.
AC P02701; Q91958;
DT 21-OUL-1986 (Rel. 01, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Avidin precursor.
GN AVD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87203384; PubMed=3575102;
RA Gope M.L., Keinaenen R.A., Kristo P.A., Conneely O.M., Beattie W.G.,
RA Zarucki-Schulz T., O'Malley B.W., Kulomaa M.S.;
RT "Molecular cloning of the chicken avidin cDNA.";
RL Nucleic Acids Res. 15:3595-3606(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90355928; PubMed=2143802;
RA Chandra G., Gray J.G.;
RT "Cloning and expression of avidin in Escherichia coli.";
RL Meth. Enzymol. 184:70-79(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=White Leghorn; TISSUE=Oviduct;
RX MEDLINE=95394357; PubMed=765080;
RA Wallen M.J., Laukkanen M.O., Kulomaa M.S.;
RT "Cloning and sequencing of the chicken egg-white avidin-encoding gene
RT and its relationship with the avidin-related genes Av4-Av5.";
RL Gene 161:205-209(1995).
RN [4]
RP SEQUENCE OF 25-152.
RX MEDLINE=71107558; PubMed=5100763;
RA Delange R.J., Huang T.-S.;
RT "Egg white avidin. 3. Sequence of the 78-residue middle cyanogen
RT bromide peptide. Complete amino acid sequence of the protein
RT subunit.";
RL J. Biol. Chem. 246:698-709(1971).
RN [5]
RP IMPORTANCE OF TYR IN BIOTIN-BINDING.
RX MEDLINE=90351377; PubMed=2386489;
RA Gitlin G., Bayer E.A., Wilchek M.;

```

RT "Studies on the biotin-binding sites of avidin and streptavidin.
 RT Tyrosine residues are involved in the binding site.",
 RT Biochem. J. 269:527-530 (1990).
 RL [6]
 RP BIOTIN-BINDING STUDIES.
 RX MEDLINE=91378911; PubMed=1896347;
 RA Hüller V., Bayer E.A., Milschek M.;
 RT "Studies on the biotin-binding site of avidin. Minimized fragments
 RT that bind biotin.";
 RL Biochem. J. 278:573-585 (1991).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
 RX MEDLINE=93281699; PubMed=8506353;
 RA Livnah O., Bayer E.A., Milschek M., Susman J.L.;
 RT "Three-dimensional structures of avidin and the avidin-biotin
 RT complex.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5076-5080 (1993).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
 RX MEDLINE=93294833; PubMed=8515446;
 RA Pugliese L., Coda A., Malcovati M., Bolognesi M.;
 RT "Three-dimensional structure of the tetragonal crystal form of
 RT egg-white avidin in its functional complex with biotin at 2.7-A
 RT resolution.";
 RL J. Mol. Biol. 231:698-710 (1993).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
 RX MEDLINE=98430987; PubMed=9760187;
 RA Nardone E., Rosano C., Santambrogio P., Currin F., Corti A., Magni F.,
 RA Sticcardi A.G., Paganeli G., Losso R., Aprea B., Bolognesi M.,
 RA Sidoli A., Arosio P.;
 RT "Biochemical characterization and crystal structure of a recombinant
 RT hen avidin and its acidic mutant expressed in Escherichia coli.";
 RL Eur. J. Biochem. 256:453-460 (1998).
 CC -1- FUNCTION: THE BIOLOGICAL FUNCTION OF AVIDIN IS NOT KNOWN. FORMS A
 CC STRONG NON-COVALENT SPECIFIC COMPLEX WITH BIOTIN (ONE MOLECULE OF
 CC BIOTIN PER SUBUNIT OF AVIDIN).
 CC -1- SUBUNIT: Homotetramer.
 CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN HEN OVIDUCT AND CONCENTRATED IN
 CC EGG WHITE (WHERE IT REPRESENT 0.05% OF THE TOTAL PROTEIN).
 CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
 CC -1- DATABASE: NAME=worthington enzyme manual;
 CC WWW="http://www.worthington-biochem.com/manual/AV.html".
 CC -1- DATABASE: NAME=Prozyme technical fact sheet:
 CC WWW="http://www.prozyme.com/technical/av10data.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X05343; CAA28954.1; -
 CC EMBL: L27818; AAB59733.1; -
 CC PIR: A54975; VICH.
 CC PDB: 2AVI; 1S-VUL-93.
 CC PDB: 1AVD; 3I-JAN-94.
 CC PDB: 1AVE; 3I-JAN-94.
 CC PDB: 1RAV; 15-JUL-98.
 CC PDB: 2CAM; 15-JUL-98.
 CC PDB: 1I38; 25-DEC-02.
 CC PDB: 1LDO; 06-NOV-02.
 CC PDB: 1LDO; 06-NOV-02.
 CC PDB: 1LEL; 06-NOV-02.
 CC InterPro: IPR005469; Avidin.
 CC InterPro: IPR005468; Avidin.str.
 CC Pfam: PF01382; Avidin.1.
 CC PRINTS: PR00709; AVIDIN.
 CC PROSITE: PS00577; AVIDIN.1.
 CC GlycoProtex: Signal; Biotin; 3D-structure.
 CC FT SIGNAL 1 24

FT CHAIN 25 152 AVIDIN.
 FT DISULFID 28 107 N-LINKED (GLCNAC.).
 FT CARBOHYD 41 41 BIOTIN (NON-COVALENT).
 FT BINDING 57 57 I -> T (IN APPROX. 50% OF THE CHAINS).
 FT VARIANT 58 58 G -> S (IN REF. 3).
 FT CONFLICT 22 22 E -> Q (IN REF. 2 AND 3).
 FT CONFLICT 77 77
 FT STRAND 32 36
 FT STRAND 41 44
 FT TURN 49 50
 FT STRAND 52 59
 FT STRAND 69 77
 FT HELIX 80 82
 FT STRAND 87 92
 FT STRAND 100 109
 FT STRAND 115 124
 FT HELIX 130 135
 FT STRAND 137 146
 SQ SEQUENCE 152 AA; 16769 MW; 1D55A4491D5EPD5C CRC64;
 Query Match. 15.3%; Score 143.5; DB 1; Length 152;
 Best Local Similarity 33.3%; Pred. No. 3.8e-05;
 Matches 44; Conservative 19; Mismatches 56; Indels 13; Gaps 7;
 QY 34 VSAAEAGINGTWNQSGTFIYTA-GADGALGTYESAVGNMESRYVLGRYSAPATDG 92
 DB 23 LSRKCSLTGKMTNDGSMNTTGAVNSRGEFTGYTAV-TATSNEIKESPLHGTENTIN 81
 QY 93 SGT-ALGWTVMKNNYRAHSAATTWQGV---GGAEARINTQWLTSQTTE-ANAMKS 146
 DB 82 KRTQPFPGFTVMK-----FSESTVFTGTCFIDRCKEY-LKTMVLLRSVNDIGDWMKA 136
 QY 147 TLVGHDTFTKVK 158
 DB 137 TRVGINIFTRLR 148
 RESULT 6
 ATR6_CHICK
 ID_AVR6_CHICK STANDARD; PRT; 150 AA.
 AC P56735;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Avidin-related protein 6 precursor.
 GN AVR6.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rhode Island;
 RX MEDLINE=21070478; PubMed=11167523;
 RA Ahlroth M.K., Kola E.H., Ewald D., Meebenda J., Sazanov A., Fries R.,
 RA Kulomaa M.S.;
 RT "Characterization and chromosomal localization of the chicken avidin
 RT gene family.";
 RL Anim. Genet. 31:367-375 (2000).
 CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ237658; CAB39893.1; -
 CC DR HSSP: P02701; 1RAV.
 CC InterPro: IPR005468; Avidin.str.

```

DR InterPro: IPR005469; Avidin.
DR InterPro: IPR005468; Avidin/str.
DR Pfam: PF01382; Avidin; 1.
DR PRINTS: PR00709; Avidin.
DR PROSITE: PS00577; AVIDIN; 1.
KW Biotin; signal; Multigene family.
FT SIGNAL 1 24
FT CHAIN 1 24
FT BINDING 57 150
FT CARBOHD 54 54
FT CARBOHD 67 54
FT CARBOHD 93 93
SQ SEQUENCE 150 AA; 16468 MM; 77548380B95B1F6 CRC64;
Query Match 13.0%; Score 121.5; DB 1; Length 150;
Matches 44; Conservativity 32.8%; Pred. No. 0.0024; Indels 23; Gaps 10
OY 34 VSAEAGITGTWYNQLGSTFIYTAGA--DGALTGTYESAV---GNAEARYLTGRYDS 86
DB 23 LSAKCSLITGKMDNLGS--IMTIGAVNDGSEFNQITIAVADNCGNITRSPLLGIQHKR 80
OY 87 APAIDGSGTALGVIYAMQNNRNMAHSATTSQC-YV--GGAEARINTOMLTSITTE-AN 142
DB 81 A-----CQPTFGFIYHW--NF--SESTSVFVGQCFVDSKREKV-LKTYMLORLAVADISD 130
OY 143 AMKSTLVGHDTFTK 156
DB 131 DMKATRVGNNDFTK 144

RESULT 8
AVR7 CHICK STANDARD; PRT; 150 AA.
ID AVR7 CHICK
AC P56736; OSW6V4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Avidin-related protein 7 precursor.
GN AVR7.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rhode Island;
RX MEDLINE=21070478; PubMed=11167523;
RA Ahlroth M.K., Kola E.H., Ewald D., Masabanda J., Sazanov A., Fries R.,
RA Kulomaa M.S.;
RA "Characterization and chromosomal localization of the chicken avidin
RT gene family";
RT Anim. Genet. 31:367-375 (2000).
RL -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ237659; CAB9894.1; -.
DR HSSP; P02701; IRAY.
DR InterPro: IPR005468; Avidin/str.
DR Pfam: PF01382; Avidin; 1.
DR PROSITE: PS00577; AVIDIN; 1.
KW Biotin; signal; Multigene family.
FT SIGNAL 1 24
FT CHAIN 1 24
FT BINDING 57 150
FT CARBOHD 54 54
FT CARBOHD 67 54
FT CARBOHD 93 93
SQ SEQUENCE 150 AA; 16468 MM; 77548380B95B1F6 CRC64;
Query Match 13.0%; Score 121.5; DB 1; Length 150;
Matches 44; Conservativity 32.8%; Pred. No. 0.0024; Indels 23; Gaps 10
OY 34 VSAEAGITGTWYNQLGSTFIYTAGA--DGALTGTYESAV---GNAEARYLTGRYDS 86
DB 23 LSAKCSLITGKMDNLGS--IMTIGAVNDGSEFNQITIAVADNCGNITRSPLLGIQHKR 80
OY 87 APAIDGSGTALGVIYAMQNNRNMAHSATTSQC-YV--GGAEARINTOMLTSITTE-AN 142
DB 81 A-----CQPTFGFIYHW--NF--SESTSVFVGQCFVDSKREKV-LKTYMLORLAVADISD 130
OY 143 AMKSTLVGHDTFTK 156
DB 131 DMKATRVGNNDFTK 144

RESULT 8
AVR7 CHICK STANDARD; PRT; 150 AA.
ID AVR7 CHICK
AC P56736; OSW6V4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Avidin-related protein 7 precursor.
GN AVR7.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rhode Island;
RX MEDLINE=21070478; PubMed=11167523;
RA Ahlroth M.K., Kola E.H., Ewald D., Masabanda J., Sazanov A., Fries R.,
RA Kulomaa M.S.;
RA "Characterization and chromosomal localization of the chicken avidin
RT gene family";
RT Anim. Genet. 31:367-375 (2000).
RL -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ237659; CAB9894.1; -.
DR HSSP; P02701; IRAY.
DR InterPro: IPR005468; Avidin/str.
DR Pfam: PF01382; Avidin; 1.
DR PROSITE: PS00577; AVIDIN; 1.
KW Biotin; signal; Multigene family.
FT SIGNAL 1 24
FT CHAIN 1 24
FT BINDING 57 150
FT CARBOHD 54 54
FT CARBOHD 67 54
FT CARBOHD 93 93
SQ SEQUENCE 150 AA; 16468 MM; 77548380B95B1F6 CRC64;
Query Match 13.0%; Score 121.5; DB 1; Length 150;
Matches 44; Conservativity 32.8%; Pred. No. 0.0024; Indels 23; Gaps 10
OY 34 VSAEAGITGTWYNQLGSTFIYTAGA--DGALTGTYESAV---GNAEARYLTGRYDS 86
DB 23 LSAKCSLITGKMDNLGS--IMTIGAVNDGSEFNQITIAVADNCGNITRSPLLGIQHKR 80
OY 87 APAIDGSGTALGVIYAMQNNRNMAHSATTSQC-YV--GGAEARINTOMLTSITTE-AN 142
DB 81 A-----CQPTFGFIYHW--NF--SESTSVFVGQCFVDSKREKV-LKTYMLORLAVADISD 130
OY 143 AMKSTLVGHDTFTK 156
DB 131 DMKATRVGNNDFTK 144

RESULT 8
AVR7 CHICK STANDARD; PRT; 150 AA.
ID AVR7 CHICK
AC P56736; OSW6V4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Avidin-related protein 7 precursor.
GN AVR7.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rhode Island;
RX MEDLINE=21070478; PubMed=11167523;
RA Ahlroth M.K., Kola E.H., Ewald D., Masabanda J., Sazanov A., Fries R.,
RA Kulomaa M.S.;
RA "Characterization and chromosomal localization of the chicken avidin
RT gene family";
RT Anim. Genet. 31:367-375 (2000).
RL -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ237659; CAB9894.1; -.
DR HSSP; P02701; IRAY.
DR InterPro: IPR005468; Avidin/str.
DR Pfam: PF01382; Avidin; 1.
DR PROSITE: PS00577; AVIDIN; 1.
KW Biotin; signal; Multigene family.
FT SIGNAL 1 24
FT CHAIN 1 24
FT BINDING 57 150
FT CARBOHD 54 54
FT CARBOHD 67 54
FT CARBOHD 93 93
SQ SEQUENCE 150 AA; 16468 MM; 77548380B95B1F6 CRC64;
Query Match 13.0%; Score 121.5; DB 1; Length 150;
Matches 44; Conservativity 32.8%; Pred. No. 0.0024; Indels 23; Gaps 10
OY 34 VSAEAGITGTWYNQLGSTFIYTAGA--DGALTGTYESAV---GNAEARYLTGRYDS 86
DB 23 LSAKCSLITGKMDNLGS--IMTIGAVNDGSEFNQITIAVADNCGNITRSPLLGIQHKR 80
OY 87 APAIDGSGTALGVIYAMQNNRNMAHSATTSQC-YV--GGAEARINTOMLTSITTE-AN 142
DB 81 A-----CQPTFGFIYHW--NF--SESTSVFVGQCFVDSKREKV-LKTYMLORLAVADISD 130
OY 143 AMKSTLVGHDTFTK 156
DB 131 DMKATRVGNNDFTK 144

RESULT 8
AVR7 CHICK STANDARD; PRT; 150 AA.
ID AVR7 CHICK
AC P56736; OSW6V4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Avidin-related protein 7 precursor.
GN AVR7.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rhode Island;
RX MEDLINE=21070478; PubMed=11167523;
RA Ahlroth M.K., Kola E.H., Ewald D., Masabanda J., Sazanov A., Fries R.,
RA Kulomaa M.S.;
RA "Characterization and chromosomal localization of the chicken avidin
RT gene family";
RT Anim. Genet. 31:367-375 (2000).
RL -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ237659; CAB9894.1; -.
DR HSSP; P02701; IRAY.
DR InterPro: IPR005468; Avidin/str.
DR Pfam: PF01382; Avidin; 1.
DR PROSITE: PS00577; AVIDIN; 1.
KW Biotin; signal; Multigene family.
FT SIGNAL 1
```

CC IsoId=p10079-2; Sequence=VSP_000451;
CC -|- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND

CC DURING EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
CC LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
CC MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
CC AND ZYGOTICALLY.
CC -1- SIMILARITY: Contains 21 EGF-like domains.
CC -1- SIMILARITY: Contains 1 CUB domain.
CC -1- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR TO
CC AVIDIN/STREPTAVIDIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, L08692; AAA62164.1; -
CC EMBL, L08692; AAA62163.1; -
CC EMBL, X17530; CAA35571.1; -
CC EMBL, M17421; AAA30050.1; -
CC EMBL, X17533; CAA35573.1; -
CC PIR, A40136; A40136.
CC HSP, P01132, 1EGF.
CC DR InterPro, IPR000152; Asx hydroxyl.
CC DR InterPro, IPR005469; Avidin.
CC DR InterPro, IPR005468; Avidin/str.
CC DR InterPro, IPR000859; CUB domain.
CC DR InterPro, IPR000742; EGF_2.
CC DR InterPro, IPR001881; EGF_Ca.
CC DR InterPro, IPR001438; EGF_11.
CC DR InterPro, IPR006209; EGF like.
CC DR Pfam, PF01382; Avidin; 1.
CC DR Pfam, PF00431; CUB; 1.
CC DR Pfam, PF00008; EGF; 21.
CC DR PRINTS; PR00709; AVIDIN.
CC DR PRINTS; PR00010; EGFBLD.
CC DR SMART; SMO0042; CUB; 1.
CC DR SMART; SMO0179; EGF CA; 20.
CC DR PROSITE; PS00010; ASX HYDROXYL; 19.
CC DR PROSITE; PS00022; EGF_1; 19.
CC DR PROSITE; PS00577; AVIDIN; 1.
CC DR PROSITE; PS01180; CUB; 1.
CC DR PROSITE; PS01186; EGF_2; 19.
CC DR PROSITE; PS01187; EGF CA; 18.
CC DR Biotin; Alternative splicing; EGF-like domain; Repeat; Signal;
CC KW Glycoprotein; Calcium-binding
CC SIGNAL 1 19
CC FT CHAIN 20 1064
FT DOMAIN 20 55
FT DOMAIN 176 62
FT DOMAIN 214 175
FT DOMAIN 252 288
FT DOMAIN 290 326
FT DOMAIN 328 364
FT DOMAIN 364 402
FT DOMAIN 402 440
FT DOMAIN 440 478
FT DOMAIN 480 516
FT DOMAIN 518 554
FT DOMAIN 556 592
FT DOMAIN 594 630
FT DOMAIN 632 668
FT DOMAIN 670 706
FT DOMAIN 708 744
FT DOMAIN 746 782
FT DOMAIN 784 820
FT DOMAIN 822 858
FT DOMAIN 860 896
FT DOMAIN 898 934
FT DOMAIN 936 1064
FT DISULFID 23 34

FT DISULFID 28 43 BY SIMILARITY.
FT DISULFID 45 54 BY SIMILARITY.
FT DISULFID 180 191 BY SIMILARITY.
FT DISULFID 185 200 BY SIMILARITY.
FT DISULFID 202 211 BY SIMILARITY.
FT DISULFID 218 229 BY SIMILARITY.
FT DISULFID 223 238 BY SIMILARITY.
FT DISULFID 228 249 BY SIMILARITY.
FT DISULFID 240 256 BY SIMILARITY.
FT DISULFID 256 267 BY SIMILARITY.
FT DISULFID 261 276 BY SIMILARITY.
FT DISULFID 278 287 BY SIMILARITY.
FT DISULFID 294 305 BY SIMILARITY.
FT DISULFID 299 314 BY SIMILARITY.
FT DISULFID 316 325 BY SIMILARITY.
FT DISULFID 332 343 BY SIMILARITY.
FT DISULFID 337 352 BY SIMILARITY.
FT DISULFID 354 363 BY SIMILARITY.
FT DISULFID 370 381 BY SIMILARITY.
FT DISULFID 375 390 BY SIMILARITY.
FT DISULFID 392 401 BY SIMILARITY.
FT DISULFID 408 419 BY SIMILARITY.
FT DISULFID 413 428 BY SIMILARITY.
FT DISULFID 430 439 BY SIMILARITY.
FT DISULFID 446 457 BY SIMILARITY.
FT DISULFID 451 466 BY SIMILARITY.
FT DISULFID 468 477 BY SIMILARITY.
FT DISULFID 484 495 BY SIMILARITY.
FT DISULFID 489 504 BY SIMILARITY.
FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 522 533 BY SIMILARITY.
FT DISULFID 527 542 BY SIMILARITY.
FT DISULFID 544 553 BY SIMILARITY.
FT DISULFID 560 571 BY SIMILARITY.
FT DISULFID 565 580 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 598 609 BY SIMILARITY.
FT DISULFID 603 618 BY SIMILARITY.
FT DISULFID 620 629 BY SIMILARITY.
FT DISULFID 636 647 BY SIMILARITY.
FT DISULFID 641 656 BY SIMILARITY.
FT DISULFID 658 667 BY SIMILARITY.
FT DISULFID 674 685 BY SIMILARITY.
FT DISULFID 679 694 BY SIMILARITY.
FT DISULFID 705 723 BY SIMILARITY.
FT DISULFID 712 723 BY SIMILARITY.
FT DISULFID 717 732 BY SIMILARITY.
FT DISULFID 734 743 BY SIMILARITY.
FT DISULFID 750 761 BY SIMILARITY.
FT DISULFID 755 770 BY SIMILARITY.
FT DISULFID 772 781 BY SIMILARITY.
FT DISULFID 788 799 BY SIMILARITY.
FT DISULFID 793 808 BY SIMILARITY.
FT DISULFID 810 819 BY SIMILARITY.
FT DISULFID 826 837 BY SIMILARITY.
FT DISULFID 831 846 BY SIMILARITY.
FT DISULFID 848 857 BY SIMILARITY.
FT DISULFID 864 875 BY SIMILARITY.
FT DISULFID 869 884 BY SIMILARITY.
FT DISULFID 886 895 BY SIMILARITY.
FT DISULFID 902 913 BY SIMILARITY.
FT DISULFID 907 922 BY SIMILARITY.
FT DISULFID 924 933 BY SIMILARITY.
FT DISULFID 936 947 BY SIMILARITY.
FT CAROXYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROXYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARPELIC 477 780 Missing (in isoform IB).
FT CONFLICT 279 279 L -> S (IN REF. 2).
SQ SEQUENCE 1064 AA; 112072 MW; 2E569CA012ED6D09 CRC64;

Query Match 12.8% Score 120; DB 1; Length 1064;
Beet Local Similarity 28.4% Pred. No. 0.025; 53; Indels 24; Gaps 7;
Matches 40; Conservative 24; Mismatches 53;

QY 34 VSAAGAG--ITGTWYNOLGSTFIYTAGADGALTGTY----ESAVGNAESRYVL----TGR 83
 DB 931 VNEEEVGFCDLEGMNNECHDOVITTKTGTGMLGDYMTYNEALYAPTVVVGASANN 990
 QY 84 YDSAPATDSGTALGWTVMKNNYRNAHSAATTWSGO--YVGAGEARINTOMLTLTS--GTTEA 141
 DB 991 YDF-----PSFGFTV-----RDNGGQSTSWTGCCHLDCDEVLVYTTMINNMVSTQ 1038
 QY 142 NAMKSTLVGHDTFTKXKPSMA 162
 DB 1039 DIKSNMVGQDKWTRVEQSIA 1059

RESULT 11
 AVR2_CHICK STANDARD; PRT; 150 AA.
 ID AVP56732;
 AC 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Avidin-related protein 2 precursor.
 GN AVR2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NC NCB1_TaxID=9031;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=White leghorn; TISSUE=Oviduct;
 RX MEDLINE=94170814; PubMed=8125122;
 RA Keiminen R.A., Wallen M.J., Kristo P.A., Laukkanen M.O.,
 RA Toimela T.A., Helenius M.A., Kulomaa M.S.;
 RT "Molecular cloning and nucleotide sequence of chicken avidin-related
 RT gene 1-5."
 RL Eur. J. Biochem. 220:615-621(1994).
 CC -1- SIMILARITY: BELONGS TO THE AVIDIN/STREPTAVIDIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; 221554; -; NOT_ANNOTATED_CDS.
 CC EMBL; 221535; -; NOT_ANNOTATED_CDS.
 DR PIR; S43202; S42202.
 DR HSSP; P02701; IRAY.
 DR Interpro; IPR005469; Avidin.
 DR Interpro; IPR005468; Avidin/str.
 DR Pfam; PF01382; Avidin.1.
 DR PRINTS; PR00709; AVIDIN.
 DR PROSITE; PS00577; AVIDIN; FALSE_NEG.
 KW biotin; signal; multigene family.
 FT CHAIN 1 24 POTENTIAL.
 FT CHAIN 25 150 AVIDIN-RELATED PROTEIN 2.
 FT BINDING 57 57 BIOTIN (NON-COVALENT) (BY SIMILARITY).
 FT CARBOHYD 67 67 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 93 93 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 150 AA; 16434 MW; 763D1E2B1A93A66D CRC64;

Query Match 12.4%; Score 116.5; DB 1; Length 150;
 Best Local Similarity 32.8%; Pred. No. 0.006;
 Matches 44; Conservative 24; Mismatches 43; Indels 23; Gaps 10;

QY 34 VSAAGAGITGTWYNOLGSTFIYTAGA---DGALTGYESAV---GNASRYVLTRYYS 86
 DB 23 LSAARKSLTGEMNDLGS--IMTIGAVNDGEGDTYITAVADNPENITSLPLGIQHKR 80
 QY 87 APATDSGTALGWTVMKNNYRNAHSAATTWSGO--YV--GSAERINTOMLTLTSCTTE--AN 142

DB 81 A-----SQPTEFFYTHW--NF--SESTSVFVGQCVDSGKGV-LKTKMLQLAVDDISD 130
 QY 143 AMKSTLVGHDTFTK 156
 DB 131 DWIATRVGNDFTR 144

RESULT 12
 SLAP_CAUCR STANDARD; PRT; 1025 AA.
 ID SLAP_CAUCR
 AC P35828; Q46015; Q9RF12;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE S-layer protein (paracrystalline surface layer protein).
 GN RSAA OR CCI007.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
 OC Caulobacteraceae; Caulobacter.
 NC NCB1_TaxID=155892;
 RN (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=93007489; PubMed=1393820;
 RA Gilchrist A., Fisher J.A., Smit J.K.;
 RT "Nucleotide sequence analysis of the gene encoding the Caulobacter
 RT crescentus paracrystalline surface layer protein."
 RL Can. J. Microbiol. 38:193-202(1992).
 RN (2)
 RP REVISIONS TO 376; 636 AND 842-843.
 RA Awram P.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=J53001;
 RA Bingle W.H., Awram P.A., Nornellini J.F., Smit J.K.;
 RT "The secretion signal of C. crescentus S-layer protein is located in
 RT the C-terminal 82 amino acids of the molecule."
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN (4)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173568; PubMed=11259647;
 RA Nietman W.C., Feidbylum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Bissen J., Heidelberg J.F., Alley M.R.K., Ohka N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
 RA Kolney J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of Caulobacter crescentus."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 RN (5)
 RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=89008089; PubMed=3049545;
 RA Fisher J.A., Smit J.K., Agabian N.;
 RT "Transcriptional analysis of the major surface array gene of
 RT Caulobacter crescentus."
 RL J. Bacteriol. 170:4706-4713(1988).
 RN (6)
 RP CHARACTERIZATION.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=98292737; PubMed=9620954;
 RA Awram P., Smit J.K.;
 RT "The Caulobacter crescentus paracrystalline S-layer protein is
 RT secreted by an ABC transporter (type I) secretion apparatus."
 CC -1- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
 CC OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. PROBABLY ACTS AS A
 CC PHYSICAL BARRIER TO PARASITES AND LYTIC ENZYMES.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-

```

CC      LAYER WITH HEXAGONAL SYMMETRY. SECRETED BY AN ABC TRANSPORTER
CC      (TYPE 1) SECRETION APPARATUS.
CC      MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE
CC      SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A
CC      SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch.)
CC      -----
CC      EMBL, AF062345; AAC38665.2; -.
CC      EMBL, AF193063; AAF19365.1; -.
CC      EMBL, AE005779; AAK22991.1; ALT_INIT.
CC      PIR: A48995; A48995.
CC      HSP: P22629; ISWC.
CC      TIGR: CCI007; -.
CC      InterPro: IPR001343; Hemlysn Ca bind.
CC      Pfam: PF00353; hemolysincbind; 3.
CC      PRINTS: PR00313; CABNDNGRPT.
CC      Cell wall, S-layer; Calcium-binding; Complete proteome.
CC      INIT_MET 0
CC      SEQUENCE 1025 AA; 98001 MW; AD7A326E1363D8AC CRC64;
CC
CC      Query Match 10.7%; Score 100.5; DB 1; Length 1025;
CC      Best Local Similarity 28.8%; Pred. No. 0.92;
CC      Matches 55; Conservative 18; Mismatches 85; Indels 33; Gaps 7;
CC
QY      6 VAAIVSLTTVITASASADPSKSKAQV---SAAEAGITGTWYNQIGSTFIYTAGADG 61
DB      298 VQAAAVTALPTGVITSGIETVMVTGSAITLNTSSGVTGLTLNNTSGAQTVAAGAGQ 357
QY      62 ALTGYESAV-----GNAESRYVLTRGYDSAPLTDGSGTLGWTVMKNNYRNAHAGAT 114
DB      358 NLITATTAQAANNVAVDGGANVTVAISTG-VISGTTTGVGANSASAGTV---SVSYANST 412
QY      115 TWSGOY-VGGAERINTOWLTSGTTEANAMKSTLVGHD-----TFTVKRPSASI 164
DB      413 TTGTAIVTVGTA-----VVAQTAGAVAVNTTLQADVTVTGNSTTAIVITQTAAT 465
QY      165 DAAKAGVUNG 175
DB      466 AGATVAGRVNG 476
QY
DB

```

```

RT      "Deciphering the biology of Mycobacterium tuberculosis from the
RT      complete genome sequence."
RT      Nature 393:537-544(1998).
RT      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN= CDC 1551 / Oshkosh;
RA      Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA      Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA      Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA      Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA      Bishai W.
RT      "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT      laboratory strains."
RT      Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC      -1- CATALYTIC ACTIVITY: Acetyl-CoA + H(2)O + glyoxylate = S-malate +
CC      COA.
CC      -1- PATHWAY: Glyoxylate bypass; second step.
CC      -1- SUBUNIT: Monomer (By similarity). (By similarity).
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic
CC      -1- SIMILARITY: Belongs to the malate synthase family. GlcB subfamily.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch.)
CC      -----
CC      EMBL, Z78020; CAB01465.1; -.
CC      EMBL, AE007047; AAK46156.1; -.
CC      PIR: F70722; F70722.
CC      PDB: 1N81; 18-DEC-02.
CC      PDB: 1N8W; 18-DEC-02.
CC      TIGR: MT1885; -.
CC      Tuberculosis; Rv1837c; -.
CC      HAMAP: MF_00641; -.
CC      InterPro: IPR001465; Malate synthase.
CC      InterPro: IPR006253; Malate synthase.
CC      Pfam: PF01274; Malate synthase; 1.
CC      TIGRFAMs: TIGR01345; malate syn G; 1.
CC      Transferase; Glyoxylate bypass; Tricarboxylic acid cycle.
CC      Complete proteome; 3D-structure.
CC      ACT SITE 339 339 CATALYTIC BASE (BY SIMILARITY).
CC      FT ACT SITE 633 633 CATALYTIC ACID (BY SIMILARITY).
CC      SEQUENCE 741 AA; 80403 MW; A92F54E0FEB87C64 CRC64;
CC
CC      Query Match 10.3%; Score 96.5; DB 1; Length 741;
CC      Best Local Similarity 26.5%; Pred. No. 1.4;
CC      Matches 56; Conservative 28; Mismatches 86; Indels 41; Gaps 11;
CC
QY      9 IAVSLTTVITASASAD---PSKSKAQVSAEAGITGTWYNQIGSTFIYTAGADGALTG 65
DB      100 ITTSGVDAEITTTGPGLVVPLNARFALNAANR-WGSLYDALYGVDPVE-TDGAEGK 157
QY      66 -TYSAVGNAESRYVLTRGYDSAPATDGS-GTALGWT-----VANKNNYRNAHAGATTW 116
DB      158 PTYKRVGDKVIAVARFELDSVPLSSGSGFDATGFTVQDQVLVALPDKSTGLANQGF 217
QY      117 SGQVVGGAERINTOWLTS-----GTTEANAMKSTLV--GHDFTVKP 159
DB      218 AG-YTGAAB--PTSVLLNHGHLIELIDPESQVGTIDRAGYKDVILBSAITTIMFED 274
QY      160 SAAIDAARKA-----GVNNGNPLDAVQO 183
DB      275 SVAAVDADAVLVGYRWMLGINKGDLAAAVDK 305
QY
DB

```

RESULT 13
 MASZ_MYCTU STANDARD; PRT; 741 AA.
 ID MASZ_MYCTU
 AC Q05096;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Malate synthase G (EC 2.3.3.9).
 GN GLCB OR RV1837C OR MT1885 OR MTCY1A11.06.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98395987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekela F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltham D., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;

RESULT 14
 ID Y309 MYCGE
 AC P47551; O49317;
 DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Hypochemical lipoprotein MG309 precursor.
 GN MG309.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 ON NCBI_TaxID=2097;
 RX MEDLINE=96026346; PubMed=565993;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
 Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
 Fitchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
 Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
 Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
 Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RT "The minimal gene complement of Mycoplasma genitalium";
 RL Science 270:397-403(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=94075230; PubMed=8253680;
 RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III;
 RT "A survey of the Mycoplasma genitalium genome by using random
 sequencing";
 RL J. Bacteriol. 175:7918-7930(1993).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 (Potential).
 CC -1- SIMILARITY: BELONGS TO THE MG307 / MG309 / MG338 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U39711; AAC7153.1; -;
 DR EMBL: U02200; AAD12488.1; -;
 DR PIR: B64234; B64234.
 DR TIGR: MG309; -;
 DR PROSITE: PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Hypochemical protein; Lipoprotein; Membrane; Signal;
 KM Complete proteome.
 FT SIGNAL 1 27
 FT CHAIN 28 1225
 FT LIPID 28 28
 FT CONFLICT 1185 1185 L -> V (IN REF. 2).
 SO SEQUENCE 1225 AA; 138375 MW; C3E4B5B431B6E8 CRC64;
 Query Match 9.9%; Score 93; DB 1; Length 1225;
 Best Local Similarity 24.6%; Pred. No. 4.6;
 Matches 56; Conservative 22; Mismatches 78; Indels 72; Gaps 9;
 QY 4 IYVAAIYAVLTTYSITRASASADPS-----KDSKQVSAEAGITGTY-----NQLDS 51
 DB 12 LLLSSIAVSL--GIAVAACAQPSRTIENLFRPSSAFDDKDGSIATLYKALENREGI 68
 QY 52 TFIIVTACAGALTGYESAVGNAESRYVLGRYDAPATDGSCTALGWTAMKNYRNH 111
 DB 69 TQVLTIRLAPVLNPFEEVNDIDIKNLRTFND-----IDNS-----FVNOEQLRN-- 116
 QY 112 SATWSGQYVGGEARINTQWLTSGTTEANAKSTLVG-----HDTFK----- 156
 DB 117 -----QYRGDYVLQTDILDTNGQAN--WKLADVNKKIYDDEINKLFTKNFEYVD 168
 QY 157 -----VKPSAASIDAAKKAGVNNGNLDAVQ 182
 DB 169 KSVGVSTPLKGIENQSMNNIKIQAKFVDKXKRLRINNDAVYAIQ 216

RESULT 15
 ID STFR_ECOLI STANDARD; PRT: 1120 AA.
 AC P76072; P77560;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Side tail fiber protein homolog from lambdaoid prophage Rac.
 GN STFR OR B1372.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 ON NCBI_TaxID=562;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 Mau B., Shao Y., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RT "The complete genome sequence of Escherichia coli K-12";
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Alba A., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
 Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 Sempel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
 Takenoko K., Takeuchi Y., Wada C., Yamamoto Y., Horikuchi T.;
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map";
 RL DNA Res. 3:363-377(1996).
 CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE00234; AAC74454.1; ALT_INIT.
 DR EMBL: D90774; BAA14966.1; -;
 DR EMBL: D90775; BAA14975.1; -;
 DR PIR: G64887; G64887.
 DR Ecogene; EG13370; stfr.
 DR InterPro; IPR005003; Phage_fiber.
 DR InterPro; IPR005068; Phage_fiber_2.
 DR Pfam; PF03335; Phage_fiber_6.
 DR Pfam; PF03406; Phage_fiber_2; 1.
 KW Hypochemical protein; Fiber protein; Repeat; Complete proteome.
 SO SEQUENCE 1120 AA; 113779 MW; 54ZBS9D71EB795B4 CRC64;
 Query Match 9.8%; Score 91.5; DB 1; Length 1120;
 Best Local Similarity 24.7%; Pred. No. 5.5;
 Matches 42; Conservative 32; Mismatches 89; Indels 7; Gaps 4;
 QY 10 AVSLTIVTISASADPSKDSKQVSAEAGITGTYNQLDSFTIVTACAGALTGYES 69
 DB 219 AIVSASTATTKASEATTSARDAAASKAKSETNA--SSASASSASTATTAAGNSAKAAT 276
 QY 70 AVGNESRYVLGRYDAPATDGSCTALGWTY--AMKNYRNHSGTWSGQYVGGEARI 128
 DB 277 SETNRSSETAAG--SASAAAGSKTAAASASASTSAGOSASATTAAGSAESAASSA 334
 QY 129 NTQWLTSGTTE--ANAKSTLVGHDTFTKYVPSAASIDAAKKAGVNNGN 176
 DB 335 STATTKAGATTEQASAAARSASAAKTSETNKASETSAESSKTAASASAS 384

Tue Oct 28 09:01:56 2003

us-09-589-870b-2.rsp

Page 12

Search completed: October 27, 2003, 10:42:31
Job time : 15 secs

...

streptavidin v1 precursor - Streptomyces venezuelae
C:Species: Streptomyces venezuelae
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 10-Dec-1999
C:Accession: S57284
R:Bayar, E.A.; Kulik, T.; Adar, R.; Wilchek, M.
Biochim. Biophys. Acta 1263, 60-66, 1995
A:Title: Close similarity among streptavidin-like, biotin-binding proteins from Streptom
A:Reference number: S57284; MUID:95359204; PMID:7632734
C:Accession: S57284
A:Molecule type: DNA
A:Residues: 1-183 <BAV>
A:Experimental source: strain Tue2460
C:Superfamily: streptavidin
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin v1 #status predicted <MAT>

Query Match 99.6%; Score 932; DB 2; Length 183;
Best Local Similarity 99.5%; Pred. No. 3,7e-66;
Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKRIVAALAVSLTIVSITASADPSKDSKAQVAAEAGITGTWYNOLGSTFIYTAGD 60
Db 1 MKRIVAALAVSLTIVSITASADPSKDSKAQVAAEAGITGTWYNOLGSTFIYTAGD 60
Qy 61 GALTGTYESAVGNASRYVLTGRYPADTDSGTALGWTVMKNNYRNAHSATTWSCGY 120
Db 61 GALTGTYESAVGNASRYVLTGRYPADTDSGTALGWTVMKNNYRNAHSATTWSCGY 120
Qy 121 VGGAEARINTQWLITSGTTEANAMKSTLVGHDTFTKVPKSAASIDAARKAGVNNGNPLDA 180
Db 121 VGGTARINTQWLITSGTTEANAMKSTLVGHDTFTKVPKSAASIDAARKAGVNNGNPLDA 180
Qy 181 VQO 183
Db 181 VQO 183

RESULT 3
streptavidin v2 precursor - Streptomyces venezuelae
C:Species: Streptomyces venezuelae
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 10-Dec-1999
C:Accession: S57285
R:Bayar, E.A.; Kulik, T.; Adar, R.; Wilchek, M.
Biochim. Biophys. Acta 1263, 60-66, 1995
A:Title: Close similarity among streptavidin-like, biotin-binding proteins from Streptom
A:Reference number: S57284; MUID:95359204; PMID:7632734
C:Accession: S57285
A:Molecule type: DNA
A:Residues: 1-183 <BAV>
A:Experimental source: strain Tue2605
C:Superfamily: streptavidin
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-183/Product: streptavidin v2 #status predicted <MAT>

Query Match 95.8%; Score 897; DB 2; Length 183;
Best Local Similarity 95.1%; Pred. No. 2,1e-63;
Matches 174; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 MKRIVAALAVSLTIVSITASADPSKDSKAQVAAEAGITGTWYNOLGSTFIYTAGD 60
Db 1 MKRIVAALAVSLTIVSITASADPSKDSKAQVAAEAGITGTWYNOLGSTFIYTAGD 60
Qy 61 GALTGTYESAVGNASRYVLTGRYPADTDSGTALGWTVMKNNYRNAHSATTWSCGY 120
Db 61 GALTGTYESAVGNASRYVLTGRYPADTDSGTALGWTVMKNNYRNAHSATTWSCGY 120
Qy 121 VGGAEARINTQWLITSGTTEANAMKSTLVGHDTFTKVPKSAASIDAARKAGVNNGNPLDA 180
Db 121 VAGSEARINTQWLITSGTTEANAMKSTLVGHDTFTKVPKSAASIDAARKAGVNNGNPLDA 180
Qy 181 VQO 183
Db 181 VQO 183

Db 181 VQO 183

RESULT 4
avidin-related protein 4/5 precursor - chicken
C:Species: Gallus gallus (chicken)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-May-1999
C:Accession: S42204; S42205
R:Keinonen, R.A.; Wallen, M.J.; Kristo, P.A.; Laakkonen, M.O.; Toimela, T.A.; Helenius,
Eur. J. Biochem. 220, 615-621, 1994
A:Title: Molecular cloning and nucleotide sequence of chicken avidin-related genes 1-5.
A:Reference number: S42201; MUID:94170814; PMID:8125122
C:Accession: S42204
A:Molecule type: DNA
A:Residues: 1-150 <KEI>
A:Cross-references: EMBL:Z22883; NID:9311811
A:Experimental source: strain White Leghorn; tissue oviduct
A:Genetics: CH1
A:Accession: S42205
A:Molecule type: DNA
A:Residues: 1-150 <KEI>
A:Cross-references: EMBL:Z22882; NID:9311812
A:Experimental source: strain White Leghorn; tissue oviduct
A:Genetics: CH2
A:Genetics: <CH1>
A:Gene: avr4
A:Introns: 27/3; 96/1; 136/2
C:Genetics: <CH2>
A:Gene: avr5
A:Introns: 27/3; 96/1; 136/2
C:Superfamily: avidin
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-150/Product: avidin-related protein 4/5 #status predicted <MAT>
F:26-105/Disulfide bonds: #status predicted

Query Match 15.7%; Score 146.5; DB 2; Length 150;
Best Local Similarity 33.8%; Pred. No. 0.00011;
Matches 45; Conservative 22; Mismatches 47; Indels 19; Gaps 8;

Qy 34 VSAEAGITGTWYNOLGSTFIYTA-GADGALTGTYESAV----GNASRYVLTGRYPASAP 88
Db 23 LSARKCSLTGKWNINLDSIMTIRAVNSRGFTGTLYAVADNPENITLSPLLGIQHKRA- 81
Qy 89 ATDSSGTALGWTVMKNNYRNAHSATTWSCGYV----GGAERINTQWLITSGTTEAN-AW 144
Db 82 ----SQTFPFYVHM--NF--SESTVFTGCGCFIDRNGKEV-LKTMWLLRSSVNDISYDW 132
Qy 145 KSTLVGHDTFTKV 157
Db 133 KATRVGYNNFTRL 145

RESULT 5
VICH
avidin precursor [validated] - chicken
C:Species: Gallus gallus (chicken)
C:Date: 24-Apr-1994 #sequence_revision 04-Nov-1994 #text_change 15-Sep-2000
C:Accession: A54975; A27518; A92093; A92092; A03160
R:Wallen, M.J.; Laakkonen, M.O.; Kulomaa, M.S.
submitted to Genbank, January 1994
A:Description: Sequence of the chicken egg-white avidin gene.
A:Accession: A54975
A:Molecule type: DNA
A:Residues: 1-21, 'S', 23-152 <MAL>
A:Cross-references: GB:L27818; NID:9450255; PID:9451869
A:Experimental source: adult oviduct, strain White Leghorn
A>Note: difference at position 22 may be due to PCR error in gene sequence
R:Goep, M.L.; Keinonen, R.A.; Kristo, P.A.; Connolly, O.M.; Beattie, W.G.; Zarucki-Schu
Nucleic Acids Res. 15, 3595-3606, 1987
A:Title: Molecular cloning of the chicken avidin cDNA.
A:Reference number: A27518; MUID:87203384; PMID:3575102

A:Accession: A27518
 A:Molecule type: mRNA
 A:Residues: 1-152 <GOP>
 R:Cross-reference: GB:X05343; NID:963071; PIDN:CAA28954.1; PID:963072
 R:Delange, R.J.; Huang, T.S.
 J. Biol. Chem. 246, 698-709, 1971
 A>Title: Egg white avidin. III. Sequence of the 78-residue middle cyanogen bromide peptide
 A:Reference number: A92093; MUID:71107559; PMID:5100763
 A:Accession: A92093
 A:Molecule type: protein
 A:Residues: 25-57, 'T', 59-76, 'E', 78-152
 A:Experimental source: egg white
 A>Note: Approximately 50% of the chains have 58-116
 R:Huang, T.S.; Delange, R.J.
 J. Biol. Chem. 246, 686-697, 1971
 A>Title: Egg white avidin. II. Isolation, composition, and amino acid sequences of the
 A:Reference number: A92092; MUID:71107557; PMID:5100762
 A:Contents: sequences of tryptic peptides
 A:Accession: A92092
 A:Molecule type: protein
 A:Residues: 25-57, 'T', 59-76, 'E', 78-152
 R:Rivnah, O.; Susseman, J.
 Submitted to the Brookhaven Protein Data Bank, April 1993
 A:Reference number: A51448; PDB:2AV1
 A:Contents: annotation; X-ray crystallography, 3.0 angstroms, with biotin, residues 27-5
 R:Rivnah, O.; Bayer, E.A.; Wilchek, M.; Susseman, J.L.
 Proc. Natl. Acad. Sci. U.S.A. 90, 5076-5080, 1993
 A>Title: Three-dimensional structure of avidin and the avidin-biotin complex.
 A:Reference number: A47554; MUID:93281699; PMID:8506353
 A:Contents: annotation; X-ray crystallography, 3.0 angstroms
 R:Pugliese, L.; Code, A.; Malcovati, M.; Bolognesi, M.
 Submitted to the Brookhaven Protein Data Bank, March 1993
 A:Reference number: A51622; PDB:1AVD
 A:Contents: annotation; X-ray crystallography, 2.7 angstroms, with biotin, residues 27-5
 R:Pugliese, L.; Code, A.; Malcovati, M.; Bolognesi, M.
 Submitted to the Brookhaven Protein Data Bank, March 1993
 A:Reference number: A51623; PDB:1AVE
 A:Contents: annotation; X-ray crystallography, 2.8 angstroms, without biotin, residues 2
 J. Mol. Biol. 231, 698-710, 1993
 A>Title: Three-dimensional structure of the tetragonal crystal form of egg-white avidin
 A:Reference number: A54974; MUID:93294833; PMID:8515446
 A:Contents: annotation; X-ray crystallography, 2.7 angstroms
 C:Genetics:
 A:introns: 27/3; 98/1; 138/2
 C:Superfamily: avidin
 C:Keywords: glycoprotein
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-152/Product: avidin #status experimental <MAT>
 F:28-107/Disulfide bonds: #status experimental
 F:41/Binding site: carbohydrate (asn) (covalent) #status experimental

Query Match 15.3%; Score 143.5; DB 1; Length 152;
 Best Local Similarity 33.3%; Pred. No. 0.0002;
 Matches 44; Conservative 19; Mismatches 56; Indels 13; Gaps 7;

QY 34 VSAEAGICTGTWYNOLGSTRFTVA--GADGALTGTYSANGASRRVLTGRVDSAPATG 92
 DB 23 LSAARKSLTGKWNDDGNSMTTIGAVNSRGEFTGTITAV-TATSNKIKSPHLGTONTIN 81
 QY 93 SGT--ALGWTAVAKNNYRNASHATWSGOV--GGAEARINTOWLLTSGTTE-ANAWKS 146
 DB 82 KRQPTFGFTVNMK---FSESTTVFTGOCFIDRNGKEV-LKTMILLRSSVNDIGDMWKA 136
 QY 147 TLVGHDTFTK 158
 DB 137 TRVGINIFRLR 148

RESULT 6
 S42201
 avidin-related protein 1 precursor - chicken
 N:Alternate names: avr1

C:Species: Gallus gallus (chicken)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-May-1999
 C:Accession: S42201
 R:Keinaenen, R.A.; Wallen, M.J.; Kristo, P.A.; Laukkanen, M.O.; Toimela, T.A.; Helenius
 Eur. J. Biochem. 220, 615-621, 1994
 A>Title: Molecular cloning and nucleotide sequence of chicken avidin-related genes 1-5.
 A:Reference number: S42201; MUID:94170814; PMID:8125122
 A:Accession: S42201
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-150 <KEI>
 A:Cross-references: EMBL:221611
 C:Genetics:
 A:introns: 27/3; 96/1; 136/2
 C:Superfamily: avidin
 C:Keywords: glycoprotein
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-150/Product: avidin-related protein 1 #status predicted <MAT>
 F:28-105/Disulfide bonds: #status predicted
 F:54,67,93/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 13.0%; Score 121.5; DB 2; Length 150;
 Best Local Similarity 32.8%; Pred. No. 0.01;
 Matches 44; Conservative 24; Mismatches 43; Indels 23; Gaps 10;

QY 34 VSAEAGICTGTWYNOLGSTRFTVA--DGALTGTYSAY---GNASRYLTGRVDS 86
 DB 23 LSAARKSLTGKWNDDGNS--MTTIGAVNDGNGFNCTYTAVDNPNITRSPGLGQHKR 80
 QY 87 APATDGGGALCMWTAVAKNNYRNASHATWSGO-VY--GGAEARINTOWLLTSGTTE-AN 142
 DB 81 A-----CQPTFGFTVNM--NF--SESTSVFVQCFCFDRNGKEV-LKTMKLRVLAVDID 130

QY 143 AMKSTLVGHDTFTK 156
 DB 131 DKAKTRVGNNDPTR 144

RESULT 7
 S42203
 avidin-related protein 3 precursor - chicken
 N:Alternate names: avr3 protein
 C:Species: Gallus gallus (chicken)
 C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 07-May-1999
 C:Accession: S42203; S39800
 R:Keinaenen, R.A.; Wallen, M.J.; Kristo, P.A.; Laukkanen, M.O.; Toimela, T.A.; Helenius
 Eur. J. Biochem. 220, 615-621, 1994
 A>Title: Molecular cloning and nucleotide sequence of chicken avidin-related genes 1-5.
 A:Reference number: S42203; MUID:94170814; PMID:8125122
 A:Accession: S42203
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-150 <KEI>
 A:Cross-references: EMBL:221612; NID:965432
 R:Kunna, T.A.; Wallen, M.J.; Kulomaa, M.S.
 Biochim. Biophys. Acta 1216, 441-445, 1993
 A>Title: Induction of chicken avidin and related mRNAs after bacterial infection.
 A:Reference number: S39799; MUID:94092737; PMID:8268225
 A:Accession: S39800
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 71-150 <KUN>
 A:Cross-references: EMBL:221536; NID:965429
 C:Genetics:
 A:introns: 27/3; 96/1; 136/2
 C:Superfamily: avidin
 F:25-150/Product: avidin-related protein 3 #status predicted <MAT>
 F:28-105/Disulfide bonds: #status predicted

Query Match 12.9%; Score 120.5; DB 2; Length 150;
 Best Local Similarity 30.3%; Pred. No. 0.012;

Db 81 A-----SQPFGCTVHW--NF--SESTSVFGQCFVDRSGKEV-LKTKMQLORLAVDDISD 130

Qy 143 AKKSTLVGHDFTFK 156
| : | | : | :
Db 131 DWIATRVGNNDFTFR 144

RESULT 10

```

flagellin - Escherichia coli (strain B1 316-42)
C:Species: Escherichia coli
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 26-Aug-1999
C:Accession: A48658
R:Schoenhals, G.; Whitefield, C.
J. Bacteriol. 175, 5395-5402, 1993
A:Title: Comparative analysis of flagellin sequences from Escherichia coli strains possess
A:Reference number: A48658; MUID:93374833; PMID:8366026
A:Accession: A48658
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-595 <SC>
A:Cross-references: CB:L07389; NID:g145996; PIDN:AAA23799.1; PID:g145997
Superfamily: flagellin

```

Query Match	12.0%	Score 112	DB 2	Length 595
Best Local Similarity	26.1%	Pred. No. 0.24		
Matches 52	Conservative 30	Mismatches 87	Indels 30	Gaps 9

[illegible]

RESULT 11

Glycosylase homolog lmo2444 [imported] - *Listeria monocytogenes* (strain EGD-e)
C:Species: *Listeria monocytogenes*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AD1380
R:Glaser, F., Frisgaul, L., Buchrieser, C., Amend, A., Baquero, F., Berche, P., Bloeker
D., Dominguez-Bernal, G., Duchaud, E., Durand, L., Dussurget, O., Entliam, K.D., Feihl, H.
Science 294, 849-852, 2001
A:Authors: Krett, J., Kuhn, M., Kunst, F., Kurapkac, G., Madueno, E., Maltournam, A., Ma
ok, C., Schluerer, T., Smoes, N., Tierrier, A., Vazquez-Boland, J.A., Voss, H., Wehland,
A:Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1380
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1310 <GLA>
A:Cross-references: GB:NC_CAD00522.1; PID:ig16411932; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
:Gene: lmo2444

Query Match	10.9%;	Score 102;	DB 2;	Length 110;
Best Local Similarity	24.9%;	Pred. No. 3.4;		
Matches	48;	Conservative	23;	Mismatches 52;
				Indels 70;
				Gaps 9
Oy	12	SLTVSITASADPSKQKQVSAAEAGITGTWNQJGSLTITVYAGADGAL-TCTTYESA	70	

```

Db      860 AVTVKVPQADAD-----TVITTSASIVGTGTEME 893
Qy      71 VGN-----AESRYLTGRYDSPAPTDSCGALGWTVMAMKN-----YRNMS- 112
Db      894 TGSNPAADVADPKPVAEFTYV--DGYDK---DGAGTIIYANVXDSCGYNDLTLYKNASSD 947
Qy      113 -----ATTWSGOYVGGAEARINTQWLITSGTTEANAMKSTL---VGHDTFTKV----- 157
Db      948 NQALSIIVNNEBYKQTTLLKPNPTDMSVQSEFTLPLSAGKNSSISYKVVDTIDGKADQVSLDKV 1007
Qy      158 -----KPSASIDA 166
Db      1008 NIGFTPTVAKVEA 1020

```

RESULT 12

extracellular matrix proteinase (EC 3.4.21.-) [imported] - *Brucella melitensis* (strain
C:Species: *Brucella melitensis*
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #ext_change 01-Feb-2002
C:Accession: AE3259
R:DelVecchio, V.G.; Kapactral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
I.; Nazur, M.; Goldsmn, E.; Selkoe, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Leese,
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella meliten*
I:Reference number: AD3352; PMID:11756688

A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1374 <KUR>
A;Cross-references: GB:HE008917; PIDN:AL51240.1; PID:gi7981929; GSEPB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0058
A;Map position: 1
C;Keywords: hydrolase; serine proteinase

Query Match 10.9%; Score 102; DB 2; Length 1374;
Best Local Similarity 21.5%; Pred. No. 3.6;
Matches 45; Conservative 29; Mismatches 65; Indels 70; Gaps 7;

RESULT 13

```

flagellin - Escherichia coli (strain Su 1242)
C:Species: Escherichia coli
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 26-Aug-1999
C:Accession: B48658
R:Schoenhals, G.; Whitfield, C.
J. Bacteriol. 175, 5395-5402, 1993
A:Title: Comparative analysis of flagellin sequences from Escherichia coli strains possess
A:Reference number: A48658; MUID:93374833; PMID:836026
A:Accession: B48658
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-595 <SCH>

```

A:Cross-references: GB:L07387; NID:g290436; PIDN:AAA23797.1; PID:g290437
C:Superfamily: flagellin

Query Match 10.8%; Score 101; DB 2; Length 595;
Best Local Similarity 25.6%; Pred. No. 1.8;
Matches 51; Conservative 29; Mismatches 89; Indels 30; Gaps 9;

QY 7 AAIAVSLTTVSTITASASADPSKDSKAQV---SAAEAGITGTWYNQLGST---FIVTAGADGAL 63
DB 343 ASVTWGTTFYFKTGADAG-AATANAGVSTDTASKEVTLNKVATKQGTAAANGDTSA 401
QY 64 TGTYESAV-----GNAESRYVLTGRYDSAPAT---DGSGLAG-WTVAMK--- 104
DB 402 TTTKSGVQVQYAVPAAGDGTSAKYADNTVSNATATYTDADGEMTTIGSVTTKYSIDA 461
QY 105 NNRYNASHATWMSGQYVG--GAEARINTQWLITS-GTTEANAMKSTLVGHD---TFTKV 157
DB 462 NNGKXVTVDSCGSGKYPKVGAEVYVSANGTLTDTATSEGTVTKDPKALDEALSSIDKF 521
QY 158 KPSAASIDAKKAGVNNGN 176
DB 522 RSLGAIQNRIDLSAVTNLN 540

RESULT 14

A48995 paracrystalline surface layer protein RsaA - Caulobacter crescentus

C:Species: Caulobacter crescentus
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C:Accession: A48995
R:Glichrist, A.; Fisher, J.A.; Smit, J.
Can. J. Microbiol. 38, 193-202, 1992
A:Title: Nucleotide sequence analysis of the gene encoding the Caulobacter crescentus pA
A:Reference number: A48995; M0ID:93007489; PMID:1393820
A:Accession: A48995
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-1026 <GIL>
A:Cross-references: GB:AF062345; GB:M22663; GB:M84760; NID:g6064104; PIDN:AAK38665.2; PI
A:Experimental source: CB15A, ATCC 19089
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:116173, NCBIPI:116174)

Query Match 10.7%; Score 100.5; DB 2; Length 1026;
Best Local Similarity 28.8%; Pred. No. 3.4;
Matches 55; Conservative 18; Mismatches 85; Indels 33; Gaps 7;

QY 6 VAAIAVSLTTVSTITASASADPSKDSKAQV---SAAEAGITGTWYNQLGSTFIVTAGADG 61
DB 299 VQAAAVTALPTGVLTISIEETMNTVSGAATLNTSSGVTGLTALNTNTSGAAQVTVTAGAQ 358
QY 62 ALTGYESAV-----GNAESRYVLTGRYDSAPATDGSGLAGWTVAMKNNRYNASHAT 114
DB 359 NLTTATTAQAANNVAVDGGANVTVAATG-VTSGTTTVGANSAASGTV---SVSVANSST 413
QY 115 TWMSGQY-VGGAEARINTQWLITSGTTEANAMKSTLVGHD-----TFTKVPASASI 164
DB 414 TTTGALAVTGGTA-----VTVAQTAGNAVNTTLTGADVTVTGNSSTTAVTVQTAAAT 466
QY 165 DAAKAGVNNG 175
DB 467 AGATVAGRVNG 477

RESULT 15

C87374

S-layer protein RsaA [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: C87374

R:Nierman, W.C.; Feildlyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eissen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; M0ID:21173698; PMID:11259647

A:Accession: C87374
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1073 <STO>
A:Cross-references: GB:AE005673; NID:g13422297; PIDN:AAK22991.1; GSPDB:GN00148
C:Genetics:
A:Gene: CCI007

Query Match 10.7%; Score 100.5; DB 2; Length 1073;
Best Local Similarity 28.8%; Pred. No. 3.6;
Matches 55; Conservative 18; Mismatches 85; Indels 33; Gaps 7;

QY 6 VAAIAVSLTTVSTITASASADPSKDSKAQV---SAAEAGITGTWYNQLGSTFIVTAGADG 61
DB 346 VQAAAVTALPTGVLTISIEETMNTVSGAATLNTSSGVTGLTALNTNTSGAAQVTVTAGAQ 405
QY 62 ALTGYESAV-----GNAESRYVLTGRYDSAPATDGSGLAGWTVAMKNNRYNASHAT 114
DB 406 NLTTATTAQAANNVAVDGGANVTVAATG-VTSGTTTVGANSAASGTV---SVSVANSST 460
QY 115 TWMSGQY-VGGAEARINTQWLITSGTTEANAMKSTLVGHD-----TFTKVPASASI 164
DB 461 TTTGALAVTGGTA-----VTVAQTAGNAVNTTLTGADVTVTGNSSTTAVTVQTAAAT 513
QY 165 DAAKAGVNNG 175
DB 514 AGATVAGRVNG 524

Search completed: October 27, 2003, 10:40:36
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 27, 2003, 10:40:43 ; Search time 17 Seconds

(without alignments)
455.464 Million cell updates/sec

Title: US-09-589-870b-2

Perfect score: 936

Sequence: 1 MKRIVVAIAIVSLTVSITA.....IDAKKAGVNGNPLDAVQ 183

Scoring table: BLOSUM62

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	936	100.0	183	4	US-08-831-399-2
2	936	100.0	183	4	US-09-366-862-2
3	936	100.0	183	4	US-09-368-772-2
4	936	100.0	183	5	PCT-US93-05240-14
5	936	100.0	183	6	5168049-5
6	928	99.1	183	4	US-09-382-276-1
7	928	99.1	183	4	US-09-385-867-1
8	901.5	96.3	186	4	US-09-382-276-2
9	899.5	96.1	186	4	US-09-382-276-3
10	846	90.4	435	2	US-08-491-988-5
11	834	89.1	159	3	US-08-948-097-17
12	834	89.1	159	3	US-08-941-100-5
13	828	88.5	159	3	US-09-381-430-2
14	828	88.5	159	3	US-08-628-540-1
15	828	88.5	159	3	US-08-941-100-1
16	745	79.6	415	2	US-08-491-988-7
17	681.5	72.1	128	1	US-08-491-988-9
18	675	72.1	128	1	US-08-211-833-2
19	675	72.1	128	1	US-08-334-718-2
20	675	72.1	128	4	US-08-831-399-16
21	675	72.1	128	4	US-09-366-862-16
22	675	72.1	128	4	US-09-368-772-16
23	634	67.7	118	4	US-07-780-717C-7
24	480	51.3	128	4	US-09-285-867-4
25	143.5	15.3	152	4	US-08-831-399-4
26	143.5	15.3	152	4	US-09-366-862-4
27	143.5	15.3	152	4	US-09-368-772-4

28	110.5	11.8	892	4	US-09-336-447A-5	Sequence 5, Appl1
29	108	11.5	21	4	US-08-566-421-2	Sequence 2, Appl1
30	105.5	11.3	2736	4	US-09-352-921A-30227	Sequence 30227, A
31	103	11.0	1005	4	US-09-206-942-41	Sequence 41, Appl1
32	103	11.0	1011	4	US-09-206-942-39	Sequence 39, Appl1
33	101.5	10.8	1026	2	US-08-614-377A-7	Sequence 7, Appl1
34	101.5	10.8	1026	3	US-09-142-648B-7	Sequence 7, Appl1
35	101	10.8	20	1	US-08-211-833-3	Sequence 3, Appl1
36	101	10.8	20	1	US-08-434-718-3	Sequence 3, Appl1
37	100.5	10.7	1026	4	US-08-194-290-7	Sequence 7, Appl1
38	99.5	10.6	1036	4	US-09-206-942-73	Sequence 73, Appl1
39	99.5	10.6	1477	1	US-08-038-882-4	Sequence 4, Appl1
40	99.5	10.6	1477	1	US-08-302-832-4	Sequence 4, Appl1
41	99.5	10.6	1477	2	US-08-530-198-4	Sequence 4, Appl1
42	99.5	10.6	1477	2	US-08-469-880-4	Sequence 4, Appl1
43	99.5	10.6	1477	2	US-08-728-470-4	Sequence 4, Appl1
44	99.5	10.6	1477	2	US-08-617-697-4	Sequence 4, Appl1
45	99.5	10.6	1477	3	US-08-719-641-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-08-831-399-2
; Sequence 2, Application US/08831399
; Patent No. 6312916
; GENERAL INFORMATION:
; APPLICANT: Kopetzki, Erhard; Muller, Rainer;
; APPLICANT: Engh, Richard; Schmitt, Urban; Deger, Arno; Brandeleter, Hans
; TITLE OF INVENTION: Recombinant Inactive Core
; TITLE OF INVENTION: Streptavidin Mutants
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felte & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/831,399
; FILING DATE: 1-April-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 13 053.0
; FILING DATE: 1-April-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 37 718.8
; FILING DATE: 16-September-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6312916man D.
; REGISTRATION NUMBER: 30,946
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 183 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-831-399-2
Query Match 100.0%; Score 936; DB 4; Length 183;
Best Local Similarity 100.0%; Pred. No. 5.5e-87;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIVAAIAVSLTIVSITASASADPSKSKAQAASAEAGITGTWYNOLGSTFIYTAGAD 60
DB 1 MKRIVAAIAVSLTIVSITASASADPSKSKAQAASAEAGITGTWYNOLGSTFIYTAGAD 60
QY 61 GALTGTYSAVGNAESRYVLTGRYDSAPATDGSSTALGWTVAWKNNYRNAHSATTWSGOY 120
DB 61 GALTGTYSAVGNAESRYVLTGRYDSAPATDGSSTALGWTVAWKNNYRNAHSATTWSGOY 120
QY 121 VGGARINTOWLLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAACKAGVNNGNPLDA 180
DB 121 VGGARINTOWLLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAACKAGVNNGNPLDA 180
QY 181 VQO 183
DB 181 VQO 183

RESULT 2
US-09-366-862-2
; Sequence 2, Application US/09366862
; Patent No. 6391571
; GENERAL INFORMATION:
; APPLICANT: Kopeczki, Erhard; Muller, Rainer;
; APPLICANT: Engh, Richard; Schmitt, Urban; Deger, Arno; Brandstetter, Hans
; TITLE OF INVENTION: Recombinant Inactive Core Streptavidin Mutants
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/366,862
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/831,399
; FILING DATE: 1-April-1997
; APPLICATION NUMBER: DE 196 13 053.0
; FILING DATE: 1-April-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 37 718.8
; FILING DATE: 16-September-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6391571man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: HUBR 1105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 183 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-09-366-862-2

Query Match 100.0%; Score 936; DB 4; Length 183;
Best Local Similarity 100.0%; Pred.No.5.5e-87;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIVAAIAVSLTIVSITASASADPSKSKAQAASAEAGITGTWYNOLGSTFIYTAGAD 60
DB 1 MKRIVAAIAVSLTIVSITASASADPSKSKAQAASAEAGITGTWYNOLGSTFIYTAGAD 60
QY 61 GALTGTYSAVGNAESRYVLTGRYDSAPATDGSSTALGWTVAWKNNYRNAHSATTWSGOY 120

DB 61 GALTGTYSAVGNAESRYVLTGRYDSAPATDGSSTALGWTVAWKNNYRNAHSATTWSGOY 120
QY 121 VGGARINTOWLLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAACKAGVNNGNPLDA 180
DB 121 VGGARINTOWLLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAACKAGVNNGNPLDA 180
QY 181 VQO 183
DB 181 VQO 183

RESULT 3
US-09-368-772-2
; Sequence 2, Application US/09368772
; Patent No. 6417331
; GENERAL INFORMATION:
; APPLICANT: Kopeczki, Erhard; Muller, Rainer;
; APPLICANT: Engh, Richard; Schmitt, Urban; Deger, Arno; Brandstetter, Hans
; TITLE OF INVENTION: Recombinant Inactive Core Streptavidin Mutants
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/368,772
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/831,399
; FILING DATE: 1-April-1997
; APPLICATION NUMBER: DE 196 13 053.0
; FILING DATE: 1-April-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 196 37 718.8
; FILING DATE: 16-September-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6417331man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: HUBR 1105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 183 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-09-368-772-2

Query Match 100.0%; Score 936; DB 4; Length 183;
Best Local Similarity 100.0%; Pred.No.5.5e-87;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIVAAIAVSLTIVSITASASADPSKSKAQAASAEAGITGTWYNOLGSTFIYTAGAD 60
DB 1 MKRIVAAIAVSLTIVSITASASADPSKSKAQAASAEAGITGTWYNOLGSTFIYTAGAD 60
QY 61 GALTGTYSAVGNAESRYVLTGRYDSAPATDGSSTALGWTVAWKNNYRNAHSATTWSGOY 120
DB 61 GALTGTYSAVGNAESRYVLTGRYDSAPATDGSSTALGWTVAWKNNYRNAHSATTWSGOY 120
QY 121 VGGARINTOWLLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAACKAGVNNGNPLDA 180

Db 121 VCGAERINTQWLTSQTTEANAMKSTLVGHDTFTVKVPSAASIDAKKAGVNNGNPLDA 180
QY 181 VQQ 183
Db 181 VQQ 183

RESULT 4
PCT-US93-05240-14

; Sequence 14, Application PCT/TUS9305240
; GENERAL INFORMATION:
; APPLICANT: NAGARAJAN, VASANTHA
; TITLE OF INVENTION: PRODUCTION OF STREPTAVIDIN FROM BACILLUS
; TITLE OF INVENTION: SUBTILIS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DU PONT COMPANY
; STREET: BARLEY MILL PLAZA 36
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
; ZIP: 19880-0036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/05240
; FILING DATE: 19930527
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: GEIGER, KATHLEEN W
; REFERENCE/DOCKET NUMBER: CR 9029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-2118
; TELEFAX: 302-892-7949
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 183 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; PCT-US93-05240-14

Query Match 100.0%; Score 936; DB 5; Length 183;
Best Local Similarity 100.0%; Pred. No. 5.5e-87;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIVVAIAVSLTTVSITASADPSKDSKAQVSAEAGITGTWYNQIGSTFIYTAGAD 60
Db 1 MKRIVVAIAVSLTTVSITASADPSKDSKAQVSAEAGITGTWYNQIGSTFIYTAGAD 60
QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDSSGTLGWTVMKNNYRNAHSAATTWSGOY 120
Db 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDSSGTLGWTVMKNNYRNAHSAATTWSGOY 120
QY 121 VCGAERINTQWLTSQTTEANAMKSTLVGHDTFTVKVPSAASIDAKKAGVNNGNPLDA 180
Db 121 VCGAERINTQWLTSQTTEANAMKSTLVGHDTFTVKVPSAASIDAKKAGVNNGNPLDA 180
QY 181 VQQ 183
Db 181 VQQ 183

RESULT 5
5168049-5
; Patent No. 5168049
; APPLICANT: MEADE, HARRY M.; GARWIN, JEFFREY L.
; TITLE OF INVENTION: PRODUCTION OF STREPTAVIDIN-LIKE
; POLYPEPTIDES

; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/185,329
; FILING DATE: 21-APR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 656,873
; FILING DATE: 02-OCT-1984
; SEQ ID NO: 5:
; LENGTH: 183
5168049-5

Query Match 100.0%; Score 936; DB 6; Length 183;
Best Local Similarity 100.0%; Pred. No. 5.5e-87;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIVVAIAVSLTTVSITASADPSKDSKAQVSAEAGITGTWYNQIGSTFIYTAGAD 60
Db 1 MKRIVVAIAVSLTTVSITASADPSKDSKAQVSAEAGITGTWYNQIGSTFIYTAGAD 60
QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDSSGTLGWTVMKNNYRNAHSAATTWSGOY 120
Db 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDSSGTLGWTVMKNNYRNAHSAATTWSGOY 120
QY 121 VCGAERINTQWLTSQTTEANAMKSTLVGHDTFTVKVPSAASIDAKKAGVNNGNPLDA 180
Db 121 VCGAERINTQWLTSQTTEANAMKSTLVGHDTFTVKVPSAASIDAKKAGVNNGNPLDA 180
QY 181 VQQ 183
Db 181 VQQ 183

RESULT 6
US-09-382-276-1

; Sequence 1, Application US/09382276
; Patent No. 6413934
; GENERAL INFORMATION:
; APPLICANT: Strayton, Patrick S.
; APPLICANT: McDevitt, Todd C.
; TITLE OF INVENTION: Streptavidin Mutants Having Secondary Functional
; FILE REFERENCE: UMS 104
; CURRENT APPLICATION NUMBER: US/09/382,276
; CURRENT FILING DATE: 1999-08-25
; EARLIER APPLICATION NUMBER: 60/097,816
; EARLIER FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Streptomyces avidinii
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(183)
; OTHER INFORMATION: Wild Type Streptavidin
; FEATURE:
; NAME/KEY: PROPEP
; LOCATION: (1)..(24)
; FEATURE:
; NAME/KEY: CHAIN
; LOCATION: (25)..(183)
US-09-382-276-1

Query Match 99.1%; Score 928; DB 4; Length 183;
Best Local Similarity 99.5%; Pred. No. 3.6e-86;
Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKRIVVAIAVSLTTVSITASADPSKDSKAQVSAEAGITGTWYNQIGSTFIYTAGAD 60
Db 1 MKRIVVAIAVSLTTVSITASADPSKDSKAQVSAEAGITGTWYNQIGSTFIYTAGAD 60

QY 61 GALTGTYSAVGNABSRVLTGRYDSAPATDGGTALGWTVMKNNYNNAHSATTWSGQY 120
DB 61 GALTGTYSAVGNABSRVLTGRYDSAPATDGGTALGWTVMKNNYNNAHSATTWSGQY 120
QY 121 VGGAEARINTOWMLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAACKAGVNNGNPLDA 180
DB 121 VGGAEARINTOWMLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAACKAGVNNGNPLDA 180
QY 181 VQO 183
DB 181 VQO 183

RESULT 7
US-09-285-867-1
; Sequence 1, Application US/09285867
; Patent No. 6492492
; GENERAL INFORMATION:
; APPLICANT: Stayton, Patrick S.
; TITLE OF INVENTION: Circularly Permuted Biotin Binding Proteins
; FILE REFERENCE: UMS 103
; CURRENT APPLICATION NUMBER: US/09/285,867
; CURRENT FILING DATE: 1999-04-02
; EARLIER APPLICATION NUMBER: 60/080,560
; EARLIER FILING DATE: 1998-04-03
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Streptavidin
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(183)
US-09-285-867-1

Query Match 99.1%; Score 928; DB 4; Length 183;
Best Local Similarity 99.5%; Pred. No. 3,6e-86;
Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKRIVVAIAVSLTIVSITASASADPSKSKAQSAAEAGITGTWYNOLGSTFIYTAGAD 60
DB 1 MKRIVVAIAVSLTIVSITASASADPSKSKAQSAAEAGITGTWYNOLGSTFIYTAGAD 60
QY 61 GALTGTYSAVGNABSRVLTGRYDSAPATDGGTALGWTVMKNNYNNAHSATTWSGQY 120
DB 61 GALTGTYSAVGNABSRVLTGRYDSAPATDGGTALGWTVMKNNYNNAHSATTWSGQY 120
QY 121 VGGAEARINTOWMLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAACKAGVNNGNPLDA 180
DB 121 VGGAEARINTOWMLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAACKAGVNNGNPLDA 180
QY 181 VQO 183
DB 181 VQO 183

RESULT 8
US-09-382-276-2
; Sequence 2, Application US/09382276
; Patent No. 6413934
; GENERAL INFORMATION:
; APPLICANT: Stayton, Patrick S.
; APPLICANT: McDevitt, Todd C.
; APPLICANT: Nelson, Kjell J.
; TITLE OF INVENTION: Streptavidin Mutants Having Secondary Functional
; FILE REFERENCE: UMS 104
; CURRENT APPLICATION NUMBER: US/09/382,276
; CURRENT FILING DATE: 1999-08-25
; EARLIER APPLICATION NUMBER: 60/097,816
; EARLIER FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 9

SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: FN-SA
; NAME/KEY: PROPEP
; LOCATION: (1)..(24)
; FEATURE:
; NAME/KEY: CHAIN
; LOCATION: (25)..(186)
US-09-382-276-2

Query Match 96.3%; Score 901.5; DB 4; Length 186;
Best Local Similarity 96.2%; Pred. No. 1.7e-83;
Matches 179; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

QY 1 MKRIVVAIAVSLTIVSITASASADPSKSKAQSAAEAGITGTWYNOLGSTFIYTAGAD 60
DB 1 MKRIVVAIAVSLTIVSITASASADPSKSKAQSAAEAGITGTWYNOLGSTFIYTAGAD 60
QY 61 GALTGTYSAVGNABSRVLTGRYDSAPATDGGTALGWTVMKNNYNNAHSATTWS 117
DB 61 GALTGTYSAVGNABSRVLTGRYDSAPATDGGTALGWTVMKNNYNNAHSATTWS 120
QY 118 GGVGAEARINTOWMLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAACKAGVNNGNP 177
DB 121 GGVGAEARINTOWMLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAACKAGVNNGNP 180
QY 178 LDAVQO 183
DB 181 LDAVQO 186

RESULT 9
US-09-382-276-3
; Sequence 3, Application US/09382276
; Patent No. 6413934
; GENERAL INFORMATION:
; APPLICANT: Stayton, Patrick S.
; APPLICANT: McDevitt, Todd C.
; APPLICANT: Nelson, Kjell J.
; TITLE OF INVENTION: Streptavidin Mutants Having Secondary Functional
; FILE REFERENCE: UMS 104
; CURRENT APPLICATION NUMBER: US/09/382,276
; CURRENT FILING DATE: 1999-08-25
; EARLIER APPLICATION NUMBER: 60/097,816
; EARLIER FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: OSTP-SA
; NAME/KEY: PROPEP
; LOCATION: (1)..(24)
; FEATURE:
; NAME/KEY: CHAIN
; LOCATION: (25)..(186)
US-09-382-276-3

Query Match 96.1%; Score 899.5; DB 4; Length 186;
Best Local Similarity 96.2%; Pred. No. 2.8e-83;
Matches 179; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

QY 1 MKRIVVAIAVSLTIVSITASASADPSKSKAQSAAEAGITGTWYNOLGSTFIYTAGAD 60

Db 1 MRKIVAAIAVSLTIVSITASASADPSKQVSAEAGITGTWYNQLSGTFIVTAGD 60
QY 61 GALTGYESAIVGAESRYVLTGRYDSPA---TDGSGTALGTVAMKNNYRNAHSATTS 117
Db 61 GALTGYESAIVGAESRYVLTGRYDSPA---TDGSGTALGTVAMKNNYRNAHSATTS 120
QY 118 GQYVGGAEARINTQWLLTSGTTEANAMKSTLVGHDTFTTKVPSAASIDAAKKAGVNNGNP 177
Db 121 GQYVGGAEARINTQWLLTSGTTEANAMKSTLVGHDTFTTKVPSAASIDAAKKAGVNNGNP 180
QY 178 LDAVQO 183
Db 181 LDAVQO 186

RESULT 10
US-08-491-988-5
; Sequence 5, Application US/08491988
; Patent No. 5973116
; GENERAL INFORMATION:
; APPLICANT: EPIENETOS, AGAMENON A.
; APPLICANT: SPOONER, ROBERT A.
; APPLICANT: DEONARIN, MAHENDRA
; TITLE OF INVENTION: Compounds for targeting
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MCAULAY NISSEN GOLDBERG KIEL & HAND, LLP
; STREET: 261 MADISON AVENUE
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10016-2391
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/491,988
; FILING DATE: 18-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDBERG, JUDS E.
; REGISTRATION NUMBER: 24,408
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-986-4090
; TELEFAX: 212-818-9479
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 435 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-491-988-5

Query Match 90.4%; Score 846; DB 2; Length 435;
Best Local Similarity 98.8%; Pred. No. 2.2e-77;
Matches 162; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 20 ASASAPSKDSKAQVSAEAGITGTWYNQLSGTFIVTAGADGALTGYESAIVGAESRYV 79
Db 272 AAAPAPSKDSKAQVSAEAGITGTWYNQLSGTFIVTAGADGALTGYESAIVGAESRYV 331
QY 80 LTGRYSAPATDSSGTLGTVAMKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTT 139
Db 332 LTGRYSAPATDSSGTLGTVAMKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTT 391
QY 140 EAAMKSTLVGHDTFTTKVPSAASIDAAKKAGVNNGNPLDAVQO 183
Db 392 EAAMKSTLVGHDTFTTKVPSAASIDAAKKAGVNNGNPLDAVQO 435

RESULT 11

US-08-948-097-17
; Sequence 17, Application US/08948097C
; Patent No. 6103493
; GENERAL INFORMATION:
; APPLICANT: Skeira, Arne
; APPLICANT: Voss, Selma
; TITLE OF INVENTION: Streptavidin Muteins
; FILE REFERENCE: HUBB 119
; CURRENT APPLICATION NUMBER: US/08/948,097C
; EARLIER FILING DATE: 1997-10-09
; NUMBER OF SEQ ID NOS: 17
; LENGTH: 159
; TYPE: PRP
; ORGANISM: Streptomyces avidinii
; US-08-948-097-17

Query Match 89.1%; Score 834; DB 3; Length 159;
Best Local Similarity 100.0%; Pred. No. 9.2e-77;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 DSKDSKAQVSAEAGITGTWYNQLSGTFIVTAGADGALTGYESAIVGAESRYVLTGRY 84
Db 1 DSKDSKAQVSAEAGITGTWYNQLSGTFIVTAGADGALTGYESAIVGAESRYVLTGRY 60
QY 85 DSAPATDSSGTLGTVAMKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAM 144
Db 61 DSAPATDSSGTLGTVAMKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAM 120
QY 145 KSTLVGHDTFTTKVPSAASIDAAKKAGVNNGNPLDAVQO 183
Db 121 KSTLVGHDTFTTKVPSAASIDAAKKAGVNNGNPLDAVQO 159

RESULT 12
US-08-941-100-5
; Sequence 5, Application US/08941100B
; Patent No. 6207390
; GENERAL INFORMATION:
; APPLICANT: Cantor, Charles R.
; APPLICANT: Sano, Takeshi
; TITLE OF INVENTION: Reduced Affinity Streptavidin
; FILE REFERENCE: BU-03165
; CURRENT APPLICATION NUMBER: US/08/941,100B
; PRIOR FILING DATE: 1997-10-03
; PRIOR APPLICATION NUMBER: 08/469,353
; PRIOR FILING DATE: 1995-06-06
; PRIOR APPLICATION NUMBER: 08/420,010
; PRIOR FILING DATE: 1995-04-11
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 159
; TYPE: PRP
; ORGANISM: Streptomyces avidinii
; US-08-941-100-5

Query Match 89.1%; Score 834; DB 3; Length 159;
Best Local Similarity 100.0%; Pred. No. 9.2e-77;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 DSKDSKAQVSAEAGITGTWYNQLSGTFIVTAGADGALTGYESAIVGAESRYVLTGRY 84
Db 1 DSKDSKAQVSAEAGITGTWYNQLSGTFIVTAGADGALTGYESAIVGAESRYVLTGRY 60
QY 85 DSAPATDSSGTLGTVAMKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAM 144
Db 61 DSAPATDSSGTLGTVAMKNNYRNAHSATTWSGQYVGGAEARINTQWLLTSGTTEANAM 120
QY 145 KSTLVGHDTFTTKVPSAASIDAAKKAGVNNGNPLDAVQO 183

Db 121 KSTLVGHDTFTKVPASASIDAAKAGVNGNPLDAVQ 159

RESULT 13

US-09-381-430-2
Sequence 2, Application US/09381430

Patent No. 6368813

GENERAL INFORMATION:

APPLICANT: Reznik, Gabriel O.

APPLICANT: Sano, Takeshi

APPLICANT: Vajda, Sandor

APPLICANT: Smith, Cassandra

APPLICANT: Cantor, Charles

TITLE OF INVENTION: MULTIFLAVOR STREPTAVIDIN

FILE REFERENCE: 1586-50152

CURRENT APPLICATION NUMBER: US/09/381,430

PRIOR FILING DATE: 2000-03-23

PRIOR APPLICATION NUMBER: PCT/US98/04931

PRIOR FILING DATE: 1998-03-13

PRIOR APPLICATION NUMBER: 60/040,771

PRIOR FILING DATE: 1997-03-14

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patent in version 3.1

SEQ ID NO 2

LENGTH: 159

TYPE: PRT

ORGANISM: Streptomyces avidinii

US-09-381-430-2

Query Match 89.1%; Score 834; DB 4; Length 159;
Best Local Similarity 100.0%; Pred. No. 9.2e-77;
Matches 159; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 DSKSKAQSAAEAGITGTWNOUGSTFIYTAGADGALTGTYESAVNAESRYVLTGRY 84

DB 1 DPKSKRAQVSAEAGITGTWNOUGSTFIYTAGADGALTGTYESAVNAESRYVLTGRY 60

QY 85 DSAPATDGGTALGWTAVKNNYRNAHSATTWGGQYVGGAEARINTOMLLTSGTTEANAW 144

DB 61 DSAPATDGGTALGWTAVKNNYRNAHSATTWGGQYVGGAEARINTOMLLTSGTTEANAW 120

QY 145 KSTLVGHDTFTKVPASASIDAAKAGVNGNPLDAVQ 183

DB 121 KSTLVGHDTFTKVPASASIDAAKAGVNGNPLDAVQ 159

RESULT 14

US-08-628-540-1
Sequence 1, Application US/08628540

Patent No. 6022951

GENERAL INFORMATION:

APPLICANT: Sano, Takeshi

APPLICANT: Cantor, Charles R.

APPLICANT: Vajda, Sandor

APPLICANT: Reznik, Gabriel O.

APPLICANT: Smith, Cassandra L.

APPLICANT: Pandori, Mark W.

TITLE OF INVENTION: STREPTAVIDIN MUTANTS

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: BAKER & BOTTS, L.L.P.

STREET: 1299 Pennsylvania Avenue, N.W.

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20004-2400

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/628,540

FILING DATE: 10-APR-1996

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/420,010

FILING DATE: 11-APR-1995

APPLICATION NUMBER: 60/003,687

FILING DATE: 18-SEP-1995

ATTORNEY/AGENT INFORMATION:

NAME: Remenick, James

REGISTRATION NUMBER: 36,902

REFERENCE/DOCKET NUMBER: 016865-0244

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-639-7700

TELEFAX: 202-639-7890

TELEX:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 159 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

ORIGINAL SOURCE:

US-08-628-540-1

Query Match 88.5%; Score 828; DB 3; Length 159;
Best Local Similarity 100.0%; Pred. No. 3.7e-76;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 PSKSKAQSAAEAGITGTWNOUGSTFIYTAGADGALTGTYESAVNAESRYVLTGRYD 85

DB 2 PDKSKRAQVSAEAGITGTWNOUGSTFIYTAGADGALTGTYESAVNAESRYVLTGRYD 61

QY 86 SAPATDGGTALGWTAVKNNYRNAHSATTWGGQYVGGAEARINTOMLLTSGTTEANAWK 145

DB 62 SAPATDGGTALGWTAVKNNYRNAHSATTWGGQYVGGAEARINTOMLLTSGTTEANAWK 121

QY 146 STLVGHDFTFTKVPASASIDAAKAGVNGNPLDAVQ 183

DB 122 STLVGHDFTFTKVPASASIDAAKAGVNGNPLDAVQ 159

RESULT 15

US-08-941-100-1
Sequence 1, Application US/08941100B

Patent No. 6207390

GENERAL INFORMATION:

APPLICANT: Sano, Takeshi

APPLICANT: Cantor, Charles R.

TITLE OF INVENTION: Reduced Affinity Streptavidin

FILE REFERENCE: BU-03165

CURRENT APPLICATION NUMBER: US/08/941,100B

PRIOR FILING DATE: 1997-10-03

PRIOR APPLICATION NUMBER: 08/469,353

PRIOR FILING DATE: 1995-06-06

PRIOR APPLICATION NUMBER: 08/420,010

PRIOR FILING DATE: 1995-04-11

NUMBER OF SEQ ID NOS: 5

SOFTWARE: Patent in Ver. 2.0

SEQ ID NO 1

LENGTH: 159

TYPE: PRT

ORGANISM: Streptomyces avidinii

FEATURE:

NAME/KEY: UNSURE

LOCATION: (1)

OTHER INFORMATION: The residue in this position can be any amino

OTHER INFORMATION: acid.

US-08-941-100-1


```

Query Match      88.5%; Score 828; DB 3; Length 159;
Best Local Similarity 100.0%; Pred. No. 3.7e-76;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      26  PSKDSKAQVSAAEAGITGTWYNOLGSTFIITAGADGALTGTYSAVGNMESRYVLTGRYD 85
      |||
      2  PSKDSKAQVSAAEAGITGTWYNOLGSTFIITAGADGALTGTYSAVGNMESRYVLTGRYD 61

QY      86  SAPATDGGTALGWTYAKKNYRNRAHSATTWSGOYVGAEARINTOMLTLTSGTTEANAMK 145
      |||
      62  SAPATDGGTALGWTYAKKNYRNRAHSATTWSGOYVGAEARINTOMLTLTSGTTEANAMK 121

QY      146  STLVGHDFTTKVKPSAASIDAKKAGYVNGNPLDAVQO 183
      |||
      122  STLVGHDFTTKVKPSAASIDAKKAGYVNGNPLDAVQO 159

Db

```

Search completed: October 27, 2003, 10:44:23
 Job time : 19 secs

THIS PAGE BLANK (USPTO)

Db 121 VCGAERINTQWLLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDA 180
QY 181 VQO 183
181 VQO 183

RESULT 2

US-10-013-173-2
; Sequence 2, Application US/10013173
; Publication No. US20030095977A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen C.
; APPLICANT: Graves, Scott Stoll
; APPLICANT: Schultz, Joanne Elaine
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James A.
; APPLICANT: Reno, John M.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; FILE REFERENCE: 690022.547C1
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Streptomyces avidinii
US-10-013-173-2

Query Match 100.0%; Score 936; DB 15; Length 183;
Best Local Similarity 100.0%; Pred. No. 1.7e-85;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIVAAIAVSLTIVTSITASADPSKSKAQAQVSAEAGITGTWYNOLGSTFIYTAGAD 60
Db 1 MKRIVAAIAVSLTIVTSITASADPSKSKAQAQVSAEAGITGTWYNOLGSTFIYTAGAD 60
QY 61 GALTGYESAVGNAESRYVLTGRYDAPATDGGTALGWTVMKNNYRNAHSATTWSGOY 120
Db 61 GALTGYESAVGNAESRYVLTGRYDAPATDGGTALGWTVMKNNYRNAHSATTWSGOY 120
QY 121 VCGAERINTQWLLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDA 180
Db 121 VCGAERINTQWLLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDA 180
QY 181 VQO 183
Db 181 VQO 183

RESULT 3

US-10-150-762-2
; Sequence 2, Application US/10150762
; Publication No. US20030103948A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen C.
; APPLICANT: Graves, Scott S.
; APPLICANT: Schultz, Joanne E.
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James A.
; APPLICANT: Reno, John M.
; APPLICANT: Dearesty, Erica A.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; FILE REFERENCE: 690022.547C2
; CURRENT FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

; LENGTH: 183
; TYPE: PRT
; ORGANISM: Streptomyces avidinii
US-10-150-762-2

Query Match 100.0%; Score 936; DB 15; Length 183;
Best Local Similarity 100.0%; Pred. No. 1.7e-85;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIVAAIAVSLTIVTSITASADPSKSKAQAQVSAEAGITGTWYNOLGSTFIYTAGAD 60
Db 1 MKRIVAAIAVSLTIVTSITASADPSKSKAQAQVSAEAGITGTWYNOLGSTFIYTAGAD 60
QY 61 GALTGYESAVGNAESRYVLTGRYDAPATDGGTALGWTVMKNNYRNAHSATTWSGOY 120
Db 61 GALTGYESAVGNAESRYVLTGRYDAPATDGGTALGWTVMKNNYRNAHSATTWSGOY 120
QY 121 VCGAERINTQWLLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDA 180
Db 121 VCGAERINTQWLLTSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDA 180
QY 181 VQO 183
Db 181 VQO 183

RESULT 4

US-10-244-821-6
; Sequence 6, Application US/10244821
; Publication No. US20030143233A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen Charles
; APPLICANT: Graves, Scott Stoll
; APPLICANT: Schultz, Joanne Elaine
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James Allen
; APPLICANT: Reno, John M.
; APPLICANT: Dearesty, Erica A.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; FILE REFERENCE: 690022.547C3
; CURRENT FILING DATE: 2002-09-16
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Predicted amino acid sequence of B9E9 single chain
; OTHER INFORMATION: antibody-genomic streptavidin fusion
US-10-244-821-6

Query Match 91.4%; Score 855.5; DB 12; Length 412;
Best Local Similarity 96.5%; Pred. No. 5.1e-77;
Matches 165; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 14 TTVSITR-SASADPSKSKAQAQVSAEAGITGTWYNOLGSTFIYTAGADGALTGYESAVG 72
Db 242 TTVVSSGSGSADPSKSKAQAQVSAEAGITGTWYNOLGSTFIYTAGADGALTGYESAVG 301
QY 73 NAESRYVLTGRYDAPATDGGTALGWTVMKNNYRNAHSATTWSGOYVCGAERINTQW 132
Db 302 NAESRYVLTGRYDAPATDGGTALGWTVMKNNYRNAHSATTWSGOYVCGAERINTQW 361
QY 133 LITSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDAVQO 183
Db 362 LITSGTTEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDAVQO 412

RESULT 5

US-10-013-173-6

```
/ Sequence 6, Application US/10013173
/ Publication No. US20030095977A1
/ GENERAL INFORMATION:
/ APPLICANT: Goshorn, Stephen C.
/ APPLICANT: Graves, Scott Stoll
/ APPLICANT: Schultz, Joanne Elaine
/ APPLICANT: Lin, Yukang
/ APPLICANT: Sanderson, James A.
/ APPLICANT: Reno, John M.
/ TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
/ FILE OF INVENTION: METHODS OF USE THEREOF
/ FILE REFERENCE: 690022.547C1
/ CURRENT APPLICATION NUMBER: US/10/013.173
/ CURRENT FILING DATE: 2001-12-07
/ NUMBER OF SEQ ID NOS: 69
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 6
/ LENGTH: 412
/ TYPE: PRN
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Predicted amino acid sequence of B9E9 single chain
US-10-013-173-6

Query Match          91.4%; Score 855.5; DB 15; Length 412;
Best Local Similarity 96.5%; Pred. No. 5.1e-77;
Matches 165; Conservative 4; Mismatches 1; Indels 1; Gaps 1;

QY 14 TTVSITA-SASADPSKDSKAQVSAEAGITGTWYNQLGSTFIYTAGADGALTGTYESAVG 72
DB 242 TTVTVSSGSGSADPSKDSKAQVSAEAGITGTWYNQLGSTFIYTAGADGALTGTYESAVG 301
QY 73 NAESRYVLTRGRYDAPATDGSCTALGWTVMKNNYRNAHSATTWSCGYVGGAEARINTOW 132
DB 302 NAESRYVLTRGRYDAPATDGSCTALGWTVMKNNYRNAHSATTWSCGYVGGAEARINTOW 361
QY 133 LLTSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAKAGVNNGNPLDAVOO 183
DB 362 LLTSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAKAGVNNGNPLDAVOO 412

RESULT 6
US-10-150-762-6
/ Sequence 6, Application US/10150762.
/ Publication No. US20030103948A1
/ GENERAL INFORMATION:
/ APPLICANT: Goshorn, Stephen C.
/ APPLICANT: Graves, Scott S.
/ APPLICANT: Schultz, Joanne E.
/ APPLICANT: Lin, Yukang
/ APPLICANT: Sanderson, James A.
/ APPLICANT: Reno, John M.
/ APPLICANT: Dearstlyne, Erica A.
/ TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
/ FILE OF INVENTION: METHODS OF USE THEREOF
/ FILE REFERENCE: 690022.547C2
/ CURRENT APPLICATION NUMBER: US/10/150.762
/ CURRENT FILING DATE: 2002-05-17
/ NUMBER OF SEQ ID NOS: 90
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 6
/ LENGTH: 412
/ TYPE: PRN
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Predicted amino acid sequence of B9E9 single chain
US-10-150-762-6

Query Match          91.4%; Score 855.5; DB 15; Length 412;
Best Local Similarity 96.5%; Pred. No. 5.1e-77;
Matches 165; Conservative 4; Mismatches 1; Indels 1; Gaps 1;
```

```
QY 14 TTVSITA-SASADPSKDSKAQVSAEAGITGTWYNQLGSTFIYTAGADGALTGTYESAVG 72
DB 242 TTVTVSSGSGSADPSKDSKAQVSAEAGITGTWYNQLGSTFIYTAGADGALTGTYESAVG 301
QY 73 NAESRYVLTRGRYDAPATDGSCTALGWTVMKNNYRNAHSATTWSCGYVGGAEARINTOW 132
DB 302 NAESRYVLTRGRYDAPATDGSCTALGWTVMKNNYRNAHSATTWSCGYVGGAEARINTOW 361
QY 133 LLTSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAKAGVNNGNPLDAVOO 183
DB 362 LLTSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAKAGVNNGNPLDAVOO 412

RESULT 7
US-10-244-821-4
/ Sequence 4, Application US/10244821
/ Publication No. US2003014323A1
/ GENERAL INFORMATION:
/ APPLICANT: Goshorn, Stephen Charles
/ APPLICANT: Graves, Scott Stoll
/ APPLICANT: Schultz, Joanne Elaine
/ APPLICANT: Lin, Yukang
/ APPLICANT: Sanderson, James Allen
/ APPLICANT: Reno, John M.
/ APPLICANT: Dearstlyne, Erica A.
/ TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
/ FILE OF INVENTION: METHODS OF USE THEREOF
/ FILE REFERENCE: 690022.547C3
/ CURRENT APPLICATION NUMBER: US/10/244.821
/ CURRENT FILING DATE: 2002-09-16
/ NUMBER OF SEQ ID NOS: 92
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 4
/ LENGTH: 431
/ TYPE: PRN
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Predicted amino acid sequence for hUNR-LU-10
US-10-244-821-4

Query Match          91.3%; Score 854.5; DB 12; Length 431;
Best Local Similarity 96.5%; Pred. No. 6.8e-77;
Matches 166; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 12 SLTVSITASASADPSKDSKAQVSAEAGITGTWYNQLGSTFIYTAGADGALTGTYESAV 71
DB 261 TLTVTS-SGSGSADPSKDSKAQVSAEAGITGTWYNQLGSTFIYTAGADGALTGTYESAV 319
QY 72 GNAESRYVLTRGRYDAPATDGSCTALGWTVMKNNYRNAHSATTWSCGYVGGAEARINTO 131
DB 330 GNAESRYVLTRGRYDAPATDGSCTALGWTVMKNNYRNAHSATTWSCGYVGGAEARINTO 379
QY 132 LLTSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAKAGVNNGNPLDAVOO 183
DB 380 LLTSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAKAGVNNGNPLDAVOO 431

RESULT 8
US-10-013-173-4
/ Sequence 4, Application US/10013173
/ Publication No. US20030095977A1
/ GENERAL INFORMATION:
/ APPLICANT: Goshorn, Stephen C.
/ APPLICANT: Graves, Scott Stoll
/ APPLICANT: Schultz, Joanne Elaine
/ APPLICANT: Lin, Yukang
/ APPLICANT: Sanderson, James A.
/ APPLICANT: Reno, John M.
/ TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
/ FILE OF INVENTION: METHODS OF USE THEREOF
/ FILE REFERENCE: 690022.547C1
```

[illegible]

US/10/013,173-8	90.5%;	Score 847;	DB 15;	Length 423;
Query Match				

```
Best Local Similarity 98.2%; Pred. No. 3.7e-76;
Matches 162; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 19 TASASADPSKDSKAQVSAEAGITGTWYNOLGSTFIVTAGADALGTYESAVGNAESRY 78
:
DB 259 SSGSADPSKDSKAQVSAEAGITGTWYNOLGSTFIVTAGADALGTYESAVGNAESRY 318
:
QY 79 VLTGRYDSAPATDGSCTALGWTAVAMKNNYRNHSAATTWSGOYVGAEARINTOMLLTSGT 138
:
DB 319 VLTGRYDSAPATDGSCTALGWTAVAMKNNYRNHSAATTWSGOYVGAEARINTOMLLTSGT 378
:
QY 139 TEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDAVQ 183
:
DB 379 TEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDAVQ 423
:

RESULT 12
US-10-150
: Sequence 8, Application US/10150762
: Publication No. US20030103948A1
: GENERAL INFORMATION:
: APPLICANT: Goshorn, Stephen C.
: APPLICANT: Graves, Scott S.
: APPLICANT: Schultz, Joanne E.
: APPLICANT: Lin, Yukang
: APPLICANT: Sanderson, James A.
: APPLICANT: Reno, John M.
: APPLICANT: Dearscyne, Erica A.
: TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
: FILE REFERENCE: 690022.547C2
: CURRENT APPLICATION NUMBER: US/10/150,762
: CURRENT FILING DATE: 2002-05-17
: NUMBER OF SEQ ID NOS: 90
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8
: LENGTH: 423
: TYPE: PR
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Predicted amino acid sequence of B9E9 single chain antibody- gene
US/10/150,762-8

Query Match 90.5%; Score 847; DB 15; Length 423;
Best Local Similarity 98.2%; Pred. No. 3.7e-76;
Matches 162; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 19 TASASADPSKDSKAQVSAEAGITGTWYNOLGSTFIVTAGADALGTYESAVGNAESRY 78
:
DB 259 SSGSADPSKDSKAQVSAEAGITGTWYNOLGSTFIVTAGADALGTYESAVGNAESRY 318
:
QY 79 VLTGRYDSAPATDGSCTALGWTAVAMKNNYRNHSAATTWSGOYVGAEARINTOMLLTSGT 138
:
DB 319 VLTGRYDSAPATDGSCTALGWTAVAMKNNYRNHSAATTWSGOYVGAEARINTOMLLTSGT 378
:
QY 139 TEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDAVQ 183
:
DB 379 TEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDAVQ 423
:

RESULT 13
US-10-244-821-88
: Sequence 88, Application US/10244821
: Publication No. US2003014323A1
: GENERAL INFORMATION:
: APPLICANT: Goshorn, Stephen Charles
: APPLICANT: Graves, Scott Stoll
: APPLICANT: Schultz, Joanne Elaine
: APPLICANT: Lin, Yukang
: APPLICANT: Sanderson, James Allen
: APPLICANT: Reno, John M.
: APPLICANT: Dearscyne, Erica A.
: TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
```

```
: TITLE OF INVENTION: METHODS OF USE THEREOF
: FILE REFERENCE: 690022.547C3
: CURRENT APPLICATION NUMBER: US/10/244,821
: CURRENT FILING DATE: 2002-09-16
: NUMBER OF SEQ ID NOS: 92
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 88
: LENGTH: 438
: TYPE: PR
: ORGANISM: Mus musculus
US-10-244-821-88

Query Match 90.5%; Score 847; DB 12; Length 438;
Best Local Similarity 98.2%; Pred. No. 3.9e-76;
Matches 162; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 19 TASASADPSKDSKAQVSAEAGITGTWYNOLGSTFIVTAGADALGTYESAVGNAESRY 78
:
DB 274 SSGSADPSKDSKAQVSAEAGITGTWYNOLGSTFIVTAGADALGTYESAVGNAESRY 333
:
QY 79 VLTGRYDSAPATDGSCTALGWTAVAMKNNYRNHSAATTWSGOYVGAEARINTOMLLTSGT 138
:
DB 334 VLTGRYDSAPATDGSCTALGWTAVAMKNNYRNHSAATTWSGOYVGAEARINTOMLLTSGT 393
:
QY 139 TEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDAVQ 183
:
DB 394 TEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDAVQ 438
:

RESULT 14
US-10-244-821-49
: Sequence 49, Application US/10244821
: Publication No. US2003014323A1
: GENERAL INFORMATION:
: APPLICANT: Goshorn, Stephen Charles
: APPLICANT: Graves, Scott Stoll
: APPLICANT: Schultz, Joanne Elaine
: APPLICANT: Lin, Yukang
: APPLICANT: Sanderson, James Allen
: APPLICANT: Reno, John M.
: APPLICANT: Dearscyne, Erica A.
: TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
: FILE REFERENCE: 690022.547C3
: CURRENT APPLICATION NUMBER: US/10/244,821
: CURRENT FILING DATE: 2002-09-16
: NUMBER OF SEQ ID NOS: 92
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 49
: LENGTH: 444
: TYPE: PR
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Predicted amino acid sequence for the CC49 single
: OTHER INFORMATION: chain antibody-genomic streptavidin fusion
US-10-244-821-49

Query Match 90.5%; Score 847; DB 12; Length 444;
Best Local Similarity 98.2%; Pred. No. 4e-76;
Matches 162; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 19 TASASADPSKDSKAQVSAEAGITGTWYNOLGSTFIVTAGADALGTYESAVGNAESRY 78
:
DB 280 SSGSADPSKDSKAQVSAEAGITGTWYNOLGSTFIVTAGADALGTYESAVGNAESRY 339
:
QY 79 VLTGRYDSAPATDGSCTALGWTAVAMKNNYRNHSAATTWSGOYVGAEARINTOMLLTSGT 138
:
DB 340 VLTGRYDSAPATDGSCTALGWTAVAMKNNYRNHSAATTWSGOYVGAEARINTOMLLTSGT 399
:
QY 139 TEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDAVQ 183
:
DB 400 TEANAMKSTLVGHDTFTKVPKPSAASIDAAKAGVNNGNPLDAVQ 444
:
```

```
RESULT 15
US-10-013-173-49
; Sequence 49, Application US/10013173
; Publication No. US20030095977A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen C.
; APPLICANT: Graves, Scott Stoll
; APPLICANT: Schultz, Joanne Elaine
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James A.
; APPLICANT: Reno, Jonh M.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 690022.547C1
; CURRENT APPLICATION NUMBER: US/10/013,173
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Predicted amino acid sequence for the CC49 single
; OTHER INFORMATION: chain antibody-genomic streptavidin fusion
; OTHER INFORMATION: sequence
US-10-013-173-49

Query Match          90.5%; Score 847; DB 15; Length 444;
Best Local Similarity 98.2%; Pred. No. 4e-76;
Matches 162; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 19 TASASADPSKDSKAQVSAEAGITGTWYNQSGSTFIVTAGADGALTGTYESAVGNASERY 78
Db 280 SGSSADPSKDSKAQVSAEAGITGTWYNQSGSTFIVTAGADGALTGTYESAVGNASERY 339
QY 79 VLTGRYSAPATDSSGTALGWTVMKNNRNNAHSATTWSGOYVGAEARINTOMLLTSGT 138
Db 340 VLTGRYSAPATDSSGTALGWTVMKNNRNNAHSATTWSGOYVGAEARINTOMLLTSGT 399
QY 139 TEANAMKSTLVGHDTFTKVKPSAASIDAKKAGVNNGNPLDAVQO 183
Db 400 TEANAMKSTLVGHDTFTKVKPSAASIDAKKAGVNNGNPLDAVQO 444
```

Search completed: October 27, 2003, 10:51:47
Job time : 30 secs

PT streptavidin-like polypeptide, also joined to another protein,
 PT e.g. tissue plasminogen activator
 XX
 PS Disclosure: Fig. 2; 54pp; English.
 XX
 CC The inventors claim the DNA sequence in SA307 which codes for a
 CC streptavidin-like polypeptide (see AAN60626), and the polypeptide
 CC encoded by it (AAP60625). They also claim hybrid Sqs comprising AAN60626
 CC and a second sequence coding for another protein, polypeptide,
 CC peptide or AA (pref. tissue plasminogen activator (TPA)).
 CC (updated on 25-MAR-2003 to correct PA field.)
 CC
 XX
 SQ Sequence 183 AA;
 Query Match 100.0%; Score 936; DB 7; Length 183;
 Best Local Similarity 100.0%; Pred. No. 9.5e-76;
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKRIVAAIAVSLTTSITASADPSKDSKAQVSAEAGITGTWYNQSGTFTVTAGAD 60
 DB 1 MKRIVAAIAVSLTTSITASADPSKDSKAQVSAEAGITGTWYNQSGTFTVTAGAD 60
 QY 61 GALTGTYESAVGNABSRVLTGRYDSAPATDGGTALGWTWAKNNYRNAHSATTWSGQY 120
 DB 61 GALTGTYESAVGNABSRVLTGRYDSAPATDGGTALGWTWAKNNYRNAHSATTWSGQY 120
 QY 121 VGGAERINTQWLITSGTTEANAMKSTLVGHDTFTKVPASASIDAAKAGVNNGNPLDA 180
 DB 121 VGGAERINTQWLITSGTTEANAMKSTLVGHDTFTKVPASASIDAAKAGVNNGNPLDA 180
 QY 181 VQO 183
 DB 181 VQO 183
 RESULT 2
 AAP93530
 ID AAP93530 standard; protein; 183 AA.
 XX
 AC AAP93530;
 XX
 DT 04-JUN-1990 (first entry)
 XX
 DE Streptavidin protein.
 XX
 KM Streptavidin; Streptomyces avidinii; biotin.
 XX
 OS Streptomyces avidinii.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..24
 FT Protein /note="Leader sequence"
 FT 25..159
 FT /note="this sequence was as the basis for the design
 FT for the synthetic gene of the present invention."
 FT
 FT
 PN WO8903422-A.
 PN
 PD 20-APR-1989.
 XX
 XX
 PF 07-OCT-1988; 88WO-GB00831.
 XX
 XX
 PR 08-OCT-1987; 87GB-0023661.
 XX
 XX
 PA (BRBI-) BRIT BIO-TECHN LTD.
 XX
 PI Edwards RM;
 XX
 WPI; 1989-130040/17.
 DNA sequence encoding streptavidin and vector -
 PT comprising hybrid gene encoding fusion protein with
 PT biotin-binding activity

XX
 PS Fig 1; page 1/5; 22pp; English.
 XX
 CC Streptavidin is a 60KD protein isolated from Streptomyces avidinii that
 CC binds extremely tightly to the vitamin biotin. It is composed of four
 CC identical subunits of 15kd and binds 4 mole of biotin per mole of
 CC protein. It is structurally related to the protein avidin. It can be
 CC readily conjugated to a range of other proteins. In order to facilitate
 CC the incorporation of streptavidin into expression vectors and the
 CC production of novel chimeric proteins containing streptavidin
 CC functionality, an improved novel synthetic gene for streptavidin has
 CC been constructed (AAN90755) based on the amino acid sequence of mature
 CC streptavidin.
 CC
 XX
 SQ Sequence 183 AA;
 Query Match 100.0%; Score 936; DB 10; Length 183;
 Best Local Similarity 100.0%; Pred. No. 9.5e-76;
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKRIVAAIAVSLTTSITASADPSKDSKAQVSAEAGITGTWYNQSGTFTVTAGAD 60
 DB 1 MKRIVAAIAVSLTTSITASADPSKDSKAQVSAEAGITGTWYNQSGTFTVTAGAD 60
 QY 61 GALTGTYESAVGNABSRVLTGRYDSAPATDGGTALGWTWAKNNYRNAHSATTWSGQY 120
 DB 61 GALTGTYESAVGNABSRVLTGRYDSAPATDGGTALGWTWAKNNYRNAHSATTWSGQY 120
 QY 121 VGGAERINTQWLITSGTTEANAMKSTLVGHDTFTKVPASASIDAAKAGVNNGNPLDA 180
 DB 121 VGGAERINTQWLITSGTTEANAMKSTLVGHDTFTKVPASASIDAAKAGVNNGNPLDA 180
 QY 181 VQO 183
 DB 181 VQO 183
 RESULT 3
 AAR44491
 ID AAR44491 standard; Protein; 183 AA.
 XX
 AC AAR44491;
 XX
 DT 25-MAR-2003 (updated)
 XX
 DT 27-JUN-1994 (first entry)
 XX
 DE Streptavidin gene.
 XX
 KM Streptavidin; protein secretion; Bacillus subtilis.
 XX
 OS Streptomyces avidinii.
 XX
 FH Key Location/Qualifiers
 FT Protein /label= signal_peptide
 FT Misc-difference 1..24
 FT Peptide /label= streptavidin
 FT 37..183
 FT /note= "expressed by transformed B. subtilis"
 FT
 FT
 PN WO9324631-A1.
 PN
 PD 09-DEC-1993.
 XX
 XX
 PR 27-MAY-1993; 93WO-US05240.
 XX
 XX
 PR 29-MAY-1992; 92US-0891524.
 XX
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Nagarajan V;
 XX

DR WPI; 1993-405822/50.
 DR P-PSDB; AAO53412.
 XX
 PT Streptavidin prodn. from Bacillus subtilis - using signal protein
 PT from bacterial exo-protein and expression element from Gram
 PT positive bacterial protein.
 XX
 PS Disclosure; Fig 1b; 54pp; English.
 XX
 CC Tetrameric biologically active streptavidin is produced by secretion
 CC from Bacillus subtilis transformed with a plasmid encoding the
 CC sequence.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 XX
 SQ Sequence 183 AA;
 Query Match 100.0%; Score 936; DB 14; Length 183;
 Best Local Similarity 100.0%; Pred. No. 9.5e-76;
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKRIVVAIVLSLTTSITASADPSKSKAQSAAEAGITGTWYNQLGSTFIITAGAD 60
 DB 1 MKRIVVAIVLSLTTSITASADPSKSKAQSAAEAGITGTWYNQLGSTFIITAGAD 60
 QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTWSGOY 120
 DB 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTWSGOY 120
 QY 121 VGGAERINTOMLLTSGTTEANAMKSTLVGHDTFTFKVPSAASIDAAKKAGVNNGNPLDA 180
 DB 121 VGGAERINTOMLLTSGTTEANAMKSTLVGHDTFTFKVPSAASIDAAKKAGVNNGNPLDA 180
 QY 181 VQQ 183
 DB 181 VQQ 183

RESULT 4
 AAW29306
 ID AAW29306 standard; Protein; 183 AA.
 XX
 AC AAW29306;
 XX
 DT 27-APR-1998 (first entry)
 XX
 DE Wild-type streptavidin protein.
 XX
 KW Streptavidin; biotin; anti-interference reagent; detection; mutain;
 KM avidin; non-specific binding.
 XX
 OS Unidentified.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..24
 FT /label= signal
 FT Protein 25..183
 XX
 PN DE19637718-A1.
 XX
 PD 02-OCT-1997.
 XX
 PF 16-SEP-1996; 96DE-1037718.
 XX
 PR 01-APR-1996; 96DE-1013053.
 XX
 PA (BOEP) BOEHRINGER MANNHEIM GMBH.
 XX
 PI Brandstetter H, Deger A, Engn R, Kopetzki E, Mueller R;
 PI Schmitz U;
 XX
 DR WPI; 1997-482043/45.
 DR N-PSDB; AAT73193.
 XX

PT Streptavidin and avidin muteins with reduced binding affinity for
 PT biotin - useful for reducing interference from nonspecific binding
 PT in assays
 XX
 PS Disclosure; Page 17-18; 26pp; German.
 XX
 CC This sequence represents a streptavidin which is used in a novel method
 CC of reducing interference from non-specific binding in assays. Muteins
 CC constructed from a core streptavidin or avidin sequence are selected that
 CC differ from the native polypeptide by at least one amino acid and have a
 CC binding affinity for biotin of less than 1010 1/mole. The biotin-bindable
 CC polypeptide may be present as a polymeric conjugate, e.g. with another
 CC polypeptide or protein, especially bovine serum albumin. These muteins
 CC are used as anti-interference reagents for reducing and/or avoiding
 CC nonspecific interactions in a process for detecting an analyte. In
 CC particular, they are used in assays where the streptavidin/avidin-biotin
 CC specific binding pair is involved for qualitative and/or quantitative
 CC determination of an analyte in a test sample, e.g. a heterogeneous
 CC immunoassay or a hybridisation assay. Despite having a lower binding
 CC affinity for biotin, the muteins have high immunological cross-reactivity
 CC with native streptavidin and avidin.
 CC
 XX
 SQ Sequence 183 AA;
 Query Match 100.0%; Score 936; DB 18; Length 183;
 Best Local Similarity 100.0%; Pred. No. 9.5e-76;
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKRIVVAIVLSLTTSITASADPSKSKAQSAAEAGITGTWYNQLGSTFIITAGAD 60
 DB 1 MKRIVVAIVLSLTTSITASADPSKSKAQSAAEAGITGTWYNQLGSTFIITAGAD 60
 QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTWSGOY 120
 DB 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNAHSATTWSGOY 120
 QY 121 VGGAERINTOMLLTSGTTEANAMKSTLVGHDTFTFKVPSAASIDAAKKAGVNNGNPLDA 180
 DB 121 VGGAERINTOMLLTSGTTEANAMKSTLVGHDTFTFKVPSAASIDAAKKAGVNNGNPLDA 180
 QY 181 VQQ 183
 DB 181 VQQ 183

RESULT 5
 AAW59216
 ID AAW59216 standard; Protein; 183 AA.
 XX
 AC AAW59216;
 XX
 DT 27-AUG-1998 (first entry)
 XX
 DE S. avidinii streptavidin protein.
 XX
 KW Streptavidin; ligand; binding affinity; mutant; isolation;
 KM purification; recover; immobilise.
 XX
 OS Streptomyces avidinii.
 XX
 PN EP835934-A2.
 XX
 PD 15-APR-1998.
 XX
 PF 09-OCT-1997; 97EP-0117504.
 XX
 PR 10-OCT-1996; 96DE-1041876.
 XX
 PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
 XX
 PI Skerra A, Voss S;
 XX
 DR WPI; 1998-218868/20.
 DR

DR N-PSDB; AAV34714.

Streptavidin mutants with higher binding affinity for peptide ligands - have mutation in amino acid region 44-53, used to isolate, purify or determine fusion proteins including these ligands

PS Disclosure; Page -; 21pp; German.

CC This sequence encodes a wild-type streptavidin protein isolated from
CC Streptomyces avidinii. This sequence is used to produce mutants which
CC are used in a method to assay the binding affinity of streptavidin
CC mutants. These mutants have a mutation within the amino acid (aa) region
CC 44-53 of the wild-type protein show a higher binding affinity than the
CC wild-type for peptide ligands that include the sequence of formula
CC $\text{Trip-X-His-Pro-Gln-Phe-Y-Z}$ where X = any aa; Y and Z are both Gly,
CC or Y = Glu and Z = Arg or Lys. Recombinant streptavidin mutants can
CC be used to isolate, purify and determine proteins or to
CC determine/recover substances that contain streptavidin-binding groups.
CC Such compounds may also be used to immobilise fusions on microtitre
CC plates, microbeads or sensor chips.
CC NOTE: This sequence does not appear in the specification but is used to
CC make the mutant streptavidin proteins represented in AAM59217 and
CC AAM59218.

CC Sequence 183 AA;

Query Match 100.0%; Score 936; DB 19; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.5e-76;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIIVAAIAVSLTIVSITASASADPSKDSKAQVSAEAGITGTWYNOLGSTFTVTAGAD 60
DB 1 MKRIIVAAIAVSLTIVSITASASADPSKDSKAQVSAEAGITGTWYNOLGSTFTVTAGAD 60
QY 61 GALTGTYESAVGNABSRVLTGRYDSAPATDGSCTALGWTVMKNNYNNAHSATTWSGOY 120
DB 61 GALTGTYESAVGNABSRVLTGRYDSAPATDGSCTALGWTVMKNNYNNAHSATTWSGOY 120
QY 121 VCGAERINTOMLTGSGTTEANAMKSTLVGHDTFTKVPKSAASIDAKKAGVNNGNPLDA 180
DB 121 VCGAERINTOMLTGSGTTEANAMKSTLVGHDTFTKVPKSAASIDAKKAGVNNGNPLDA 180
QY 181 VQO 183
DB 181 VQO 183

RESULT 6

AAV17868
ID AAV17868 standard; Protein; 183 AA.

AC AAV17868;

DT 20-AUG-1999 (first entry)

DE Streptococcus streptavidin.

KM Avidin; streptavidin; batroxobin; fibrinogen converting enzyme;

KW hybrid; fusion protein; sealant; surgery; reduce bleeding; fibrin.

OS Streptococcus sp.

PN MO9929838-A1.

PD 17-JUN-1999.

PF 09-DEC-1998; 98WO-US26086.

PR 09-DEC-1997; 97US-0067978.

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Cederholm-Williams SA;

XX WPI; 1999-385599/32.

DR N-PSDB; AAX80198.

PT A fibrinogen-converting enzyme fusion protein

PS Disclosure; Page 28; 35pp; English.

CC The present invention describes a fibrinogen-converting enzyme fusion
CC protein (FCE). The fusion protein is a multidomain protein comprising:
CC (a) a FCE; and (b) a first member of a binding pair that is linked to
CC the FCE chain; (1) directly by bonds utilizing the N-terminal amino
CC groups, the C-terminal carboxy groups or side-chain functionalities;
CC (ii) via a bifunctional linkage moiety linking the groups or
CC functionalities; or (iii) by the first member binding to the second
CC member of the binding pair, where the second member of the binding pair
CC is covalently attached to the first polypeptide chain. The FCE can be
CC used in a method for producing fibrin. Fibrin is useful as a sealant in
CC surgery to, e.g. reduce bleeding by sealing blood vessels, and tissues
CC that have been dissected either in surgery or through wounding. The
CC fusion protein allows for the removal of the fibrinogen converting
CC enzyme from the fibrin sealant preparation via the binding of
CC streptavidin to a biotin solid support. The present sequence represents
CC Streptococcus streptavidin as given in the present invention.

CC Sequence 183 AA;

Query Match 100.0%; Score 936; DB 20; Length 183;
Best Local Similarity 100.0%; Pred. No. 9.5e-76;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKRIIVAAIAVSLTIVSITASASADPSKDSKAQVSAEAGITGTWYNOLGSTFTVTAGAD 60
DB 1 MKRIIVAAIAVSLTIVSITASASADPSKDSKAQVSAEAGITGTWYNOLGSTFTVTAGAD 60
QY 61 GALTGTYESAVGNABSRVLTGRYDSAPATDGSCTALGWTVMKNNYNNAHSATTWSGOY 120
DB 61 GALTGTYESAVGNABSRVLTGRYDSAPATDGSCTALGWTVMKNNYNNAHSATTWSGOY 120
QY 121 VCGAERINTOMLTGSGTTEANAMKSTLVGHDTFTKVPKSAASIDAKKAGVNNGNPLDA 180
DB 121 VCGAERINTOMLTGSGTTEANAMKSTLVGHDTFTKVPKSAASIDAKKAGVNNGNPLDA 180
QY 181 VQO 183
DB 181 VQO 183

RESULT 7

AAV44701
ID AAV44701 standard; Protein; 183 AA.

AC AAV44701;

DT 25-APR-2000 (first entry)

DE Streptavidin protein for recombinant PART27 vector.

KM Potato proteinase inhibitor-II; PPI-II; streptavidin; worm;

KW insect; plant-noxious protein; pest resistance; moth; insect; weevil;

KW grub; beetle; fly; thrip; locust; cricket; borer; mite; looper;

OS Unidentified

PN Key Location/Qualifiers

FT Peptide 1..24 /note="Signal peptide"

PD WO200004049-A1.

XX 27-JAN-2000.

PF 15-JUL-1999; 99MO-N200110.
 XX
 PR 15-JUL-1998; 98NZ-0331082.
 XX
 PA (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
 XX
 PI Christeller JT, Sutherland PW, Murray C, Markwick NP, Philip BA;
 PI Malone LA, Burgess EPJ;
 XX
 DR WPI; 2000-171244/15.
 DR N-PSDB; AAC249867.
 XX
 PT New chimeric polypeptide and composition comprising the polypeptide
 PT useful for conferring pest resistance on plants -
 XX
 PS Disclosure; Fig 12; 11pp; English.
 XX
 CC The present sequence is streptavidin, a plant-toxicous protein.
 CC Recombinant vector, PART27 expressing a chimeric polypeptide comprising
 CC streptavidin mature peptide fused to the potato proteinase inhibitor-II
 CC (PPI-II) signal peptide is targeted to the vacuole.
 CC Transformation of plant genome with the vector can produce pest
 CC resistance in plants, plant derived products and stored harvest
 CC material. Pests that can be controlled include, cotton bollworm,
 CC tropical army-worm, European corn-borer or red mite, tobacco horn worm,
 CC loopers, rice stem borer, porina, cutworms, diamondback moth, potato
 CC tuber moth, codling moth, Indian meal moth, gypsy moth, Argentine stem
 CC weevil, clover root weevil, grass-grubs, corn rootworm, rice and wheat
 CC weevils, mealworms, flour beetles, black field cricket, locusts,
 CC leafhoppers, Western flower thrips, Hessian flies or two-spotted mite.
 XX
 SQ Sequence 183 AA;
 Query Match 100.0%; Score 936; DB 21; Length 183;
 Best Local Similarity 100.0%; Pred. No. 9.5e-76;
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRKIIVAAIAVSLTTSITASASADPSKDSKAQVSAEAGITGTWYNOLGSTFIYTAGAD 60
 DB 1 MRKIIVAAIAVSLTTSITASASADPSKDSKAQVSAEAGITGTWYNOLGSTFIYTAGAD 60
 QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNHSAATTWSGOY 120
 DB 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNHSAATTWSGOY 120
 QY 121 VCGAEARINTOMLITSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAKAGVNNGNPLDA 180
 DB 121 VCGAEARINTOMLITSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAKAGVNNGNPLDA 180
 QY 181 VQQ 183
 DB 181 VQQ 183
 RESULT 8
 AAB30692 standard; Protein: 183 AA.
 ID AAB30692;
 AC AAB30692;
 XX
 DT 02-APR-2001 (first entry)
 XX
 DE Amino acid sequence of a streptavidin polypeptide.
 XX
 KM Streptavidin; tumour cell; cancer; adenocarcinoma;
 KM hematological malignancy.
 XX
 OS Streptomyces avidinii.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..24
 FT /note= "signal peptide"
 XX

PN WO200075333-A1.
 XX
 PD 14-DEC-2000.
 XX
 PF 05-JUN-2000; 2000MO-US15595.
 XX
 PR 07-JUN-1999; 99US-0137900.
 PR 03-DEC-1999; 99US-0168976.
 XX
 PA (NEOR-) NEORX CORP.
 XX
 PI Goshorn SC, Graves SS, Schultz JE, Lin Y, Sanderson JA, Reno JM;
 DR WPI; 2001-091213/10.
 DR N-PSDB; AAC6561.
 XX
 PT New vector constructs for expressing genomic streptavidin fusion
 PT proteins which are useful for targeting tumour cells associated with
 XX cancer, e.g. adenocarcinomas -
 XX
 PS Claim 18; Fig 4; 100pp; English.
 XX
 CC The present sequence represents a streptavidin polypeptide. The sequence
 CC is used to construct vectors of the invention. The specification
 CC describes vector constructs for expressing streptavidin fusion proteins.
 CC The vector comprises a first nucleic acid encoding genomic streptavidin
 CC or its functional variant operatively linked to a promoter, and a
 CC cloning site for insertion of a second nucleic acid sequence encoding a
 CC polypeptide to be fused with streptavidin. Inteposed between the
 CC promoter and the first nucleic acid sequence. Alternatively, the vector
 CC construct comprises a first nucleic acid, operatively linked to a
 CC promoter, encoding a polypeptide to be fused with streptavidin, and a
 CC cloning site for insertion of a second nucleic acid encoding at least
 CC 129 amino acids of streptavidin or its functional variant. The fusion
 CC proteins are useful for targeting tumour cells, particularly tumour cells
 CC associated with cancer, e.g. adenocarcinomas or hematological
 CC malignancies. The vector construct is useful for expressing of
 CC streptavidin fusion proteins. In particular, these are useful as tools
 CC for medical diagnostics and therapeutic purposes, e.g. for detecting the
 CC presence or absence of, or treating, a target site within a mammalian
 CC host.
 XX
 SQ Sequence 183 AA;
 Query Match 100.0%; Score 936; DB 22; Length 183;
 Best Local Similarity 100.0%; Pred. No. 9.5e-76;
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRKIIVAAIAVSLTTSITASASADPSKDSKAQVSAEAGITGTWYNOLGSTFIYTAGAD 60
 DB 1 MRKIIVAAIAVSLTTSITASASADPSKDSKAQVSAEAGITGTWYNOLGSTFIYTAGAD 60
 QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNHSAATTWSGOY 120
 DB 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGSCTALGWTVMKNNYRNHSAATTWSGOY 120
 QY 121 VCGAEARINTOMLITSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAKAGVNNGNPLDA 180
 DB 121 VCGAEARINTOMLITSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAKAGVNNGNPLDA 180
 QY 181 VQQ 183
 DB 181 VQQ 183
 RESULT 9
 AAY80512 standard; Protein: 186 AA.
 ID AAY80512;
 AC AAY80512;
 XX
 DT 06-JUN-2000 (first entry)
 XX

DE Streptomyces avidinii sps protein.
XX
KW plant somatic tissue degeneration; plant essential factor; depletion;
KW viability; sps gene; plant development; plant morphology; flower;
KW fruit plant.
XX
OS Streptomyces avidinii.
XX
PN WO200007427-A2.
XX
PD 17-FEB-2000.
XX
PF 30-JUL-1999; 99WO-IL00420.
XX
PR 03-AUG-1998; 98IL-0125632.
XX
PA (AGRI-) AGRIC RES ORG.
XX
PI Kapulnik Y, Ginzberg I;
XX
PI WPI; 2000-195402/17.
XX
DR N-PSDB; AA291073.
XX
PT Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX
PS Examples; Page 84; 91pp; English.
XX
CC The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomyces
CC avidinii sps protein as an example of a heterologous protein introduced
CC into the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 186 AA;
XX
Query Match 100.0%; Score 936; DB 21; Length 186;
Best Local Similarity 100.0%; Pred. No. 9.7e-76;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKRIVVAALAVSLTIVSITASADPSKDSKAQVSAEAGITGTWYNQIGSTFIYTAGAD 60
DB 4 MKRIVVAALAVSLTIVSITASADPSKDSKAQVSAEAGITGTWYNQIGSTFIYTAGAD 63
QY 61 GALTGTYESAVGNASRYVLTGRYDSAPATDSSGALGWTWAKNNYRAHSAATTWSQY 120
DB 64 GALTGTYESAVGNASRYVLTGRYDSAPATDSSGALGWTWAKNNYRAHSAATTWSQY 123
QY 121 VGGAERINTOWLNSGTTEANAMKSTLVGHDTFTKVPASASIDAAKKAQVNNGNPLDA 180
DB 124 VGGAERINTOWLNSGTTEANAMKSTLVGHDTFTKVPASASIDAAKKAQVNNGNPLDA 183
QY 181 VQO 183
DB 184 VQO 186
XX
RESULT 10
ID AAY80513
XX AAY80513 standard; Protein; 184 AA.
AC AAY80513;
XX
DT 06-JUN-2000 (first entry)

XX
DE Streptomyces avidinii mst protein.
XX
KW plant somatic tissue degeneration; plant essential factor; depletion;
KW viability; mst gene; plant development; plant morphology; flower;
KW fruit plant.
XX
OS Streptomyces avidinii.
XX
PN WO200007427-A2.
XX
PD 17-FEB-2000.
XX
PF 30-JUL-1999; 99WO-IL00420.
XX
PR 03-AUG-1998; 98IL-0125632.
XX
PA (AGRI-) AGRIC RES ORG.
XX
PI Kapulnik Y, Ginzberg I;
XX
PI WPI; 2000-195402/17.
XX
DR N-PSDB; AA291074.
XX
PT Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX
PS Examples; Page 85; 91pp; English.
XX
CC The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomyces
CC avidinii mst protein as an example of a heterologous protein introduced
CC into the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used
CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 184 AA;
XX
Query Match 99.5%; Score 911; DB 21; Length 184;
Best Local Similarity 100.0%; Pred. No. 2.7e-75;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 RKIIVAAALAVSLTIVSITASADPSKDSKAQVSAEAGITGTWYNQIGSTFIYTAGAD 61
DB 3 RKIIVAAALAVSLTIVSITASADPSKDSKAQVSAEAGITGTWYNQIGSTFIYTAGAD 62
QY 62 ALTGTYESAVGNASRYVLTGRYDSAPATDSSGALGWTWAKNNYRAHSAATTWSQY 121
DB 63 ALTGTYESAVGNASRYVLTGRYDSAPATDSSGALGWTWAKNNYRAHSAATTWSQY 122
QY 122 GGAERINTOWLNSGTTEANAMKSTLVGHDTFTKVPASASIDAAKKAQVNNGNPLDAV 181
DB 123 GGAERINTOWLNSGTTEANAMKSTLVGHDTFTKVPASASIDAAKKAQVNNGNPLDAV 182
QY 182 QO 183
DB 183 QO 184
XX
RESULT 11
ID AAP70492
XX AAP70492 standard; protein; 182 AA.
AC AAP70492;
XX

DT 25-MAR-2003 (updated)
 DT 06-MAR-1991 (first entry)
 XX
 DE Streptavidin sequence.
 XX
 KM Streptavidin: N-terminal; fusion gene; fusion protein;
 XX
 OS Streptomyces avidinii.
 XX
 PN MO8705026-A.
 XX
 PD 27-AUG-1987.
 XX
 PF 24-FEB-1987; 87MO-US00397.
 XX
 PR 24-FEB-1986; 86US-0833324.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PA (UNY) UNIV NEW YORK STATE.
 XX
 PI Cantor CR, Axel R, Garana C;
 XX
 DR WPI; 1987-250198/35.
 DR N-PSDB; AAN70810.
 XX
 PT DNA encoding streptavidin - obtd. by restriction endo-nuclease
 PT digestion of chromosomal DNA of Streptomyces avidinii
 XX
 PS Disclosure; Figure 3; 54pp; English.
 XX
 CC The sequence is that of streptavidin from S.avidinii. Streptavidin
 CC may be expressed from a fusion gene comprising its coding gene and DNA
 CC encoding a target protein of interest, where the streptavidin has
 CC binding sites for biotin or deriv. The streptavidin has 4 binding
 CC sites free for biotin, and is produced free of biotin
 CC contamination. Improved streptavidins may also be produced by
 CC site-directed mutagenesis. The fused gene may be used to produce
 CC labelled, chemically-modified proteins in vivo, and to isolate
 CC proteins when only the sequence of the gene is known.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 XX
 SQ Sequence 182 AA;
 XX
 Query Match 99.1%; Score 928; DB 8; Length 182;
 Best Local Similarity 99.5%; Pred. No. 4.9e-75;
 Matches 181; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 2 RKIVAAIAVSLTIVSITASASADPSKDSKAQVSAEAGITGTWYNQLGSTFIIVTAGADG 61
 DB 1 RKIVAAIAVSLTIVSITASASADPSKDSKAQVSAEAGITGTWYNQLGSTFIIVTAGADG 60
 QY 62 ALTGYESAVGNESRVLTRGRYDSAPATDGSGLALGWTVAAMKNNRNNAHSATTWSGQYV 121
 DB 61 ALTGYESAVGNESRVLTRGRYDSAPATDGSGLALGWTVAAMKNNRNNAHSATTWSGQYV 120
 QY 122 GGAERINTOWLTSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAGVNNGNPLDAV 181
 DB 121 GGAERINTOWLTSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAGVNNGNPLDAV 180
 QY 182 QQ 183
 DB 181 QQ 182

RESULT 12

AAV84020
 ID AAV84020 standard; protein; 183 AA.
 XX
 AC AAV84020;
 XX

DT 03-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of the wildtype streptavidin monomer.

XX
 KM Streptavidin: monomer; biotin binding domain; functional domain;
 KM biotin; cell adhesion peptide; adaptor; streptavidin/biotin interaction;
 KM substrate; vascular device; prosthesis.
 XX
 OS Streptomyces avidinii.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..24
 FT /notes "propeptide"
 XX
 PN MO200011152-A1.
 XX
 PD 02-MAR-2000.
 XX
 PF 25-AUG-1999; 99WO-US19481.
 XX
 PR 25-AUG-1998; 98US-0097816.
 XX
 PA (UNIW) UNIV WASHINGTON.
 XX
 PI Stayton PS, McDevitt TC, Nelson KE;
 XX
 DR WPI; 2000-224689/19.
 XX
 PT Mutant streptavidin molecule, useful as an adaptor and a coating for
 PT substrates such as vascular devices or prostheses, comprises a biotin
 PT binding domain and a secondary functional domain -
 XX
 PS Disclosure; Page 35-36; 43pp; English.
 XX
 CC The present sequence represents a wildtype streptavidin monomer. The
 CC specification describes streptavidin molecules comprising a biotin
 CC binding domain and a secondary functional domain. The molecules are
 CC adaptors with inherent effector function. They can therefore, bind to
 CC biotin and also have another function, for e.g. binding to a cell
 CC through the secondary functional domain comprising a cell adhesion
 CC peptide. The streptavidin molecules are useful as adaptors to bring,
 CC via a streptavidin/biotin interaction, the secondary functional domain
 CC into proximity with a cell or molecule to be affected and as a coating
 CC for substrates such as vascular devices or prostheses. Therefore, any
 CC compound of interest, such as a nucleic acid, protein, peptide, organic
 CC compound, inorganic compound, polysaccharide or a combination, can be
 CC targeted, delivered or immobilized using them. The molecules, when
 CC comprising an antibody for its secondary functional domain, are also
 CC useful in diagnostic applications for detecting analytes.
 XX
 SQ Sequence 183 AA;
 XX
 Query Match 99.1%; Score 928; DB 21; Length 183;
 Best Local Similarity 99.5%; Pred. No. 4.9e-75;
 Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 XX
 QY 1 MKRIVAAIAVSLTIVSITASASADPSKDSKAQVSAEAGITGTWYNQLGSTFIIVTAGAD 60
 DB 1 MKRIVAAIAVSLTIVSITASASADPSKDSKAQVSAEAGITGTWYNQLGSTFIIVTAGAD 60
 QY 61 GALTGYESAVGNESRVLTRGRYDSAPATDGSGLALGWTVAAMKNNRNNAHSATTWSGQY 120
 DB 61 GALTGYESAVGNESRVLTRGRYDSAPATDGSGLALGWTVAAMKNNRNNAHSATTWSGQY 120
 QY 121 VGGAERINTOWLTSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAGVNNGNPLDA 180
 DB 121 VGGAERINTOWLTSGTTEANAMKSTLVGHDTFTKVKPSAASIDAAGVNNGNPLDA 180
 QY 181 VQQ 183
 DB 181 VQQ 183

RESULT 13

AAM59217
 ID AAM59217 standard; protein; 183 AA.

XX	AAM59217;
AC	27-AUG-1998 (first entry)
DT	S. avidinii streptavidin mutant protein #1.
XX	
DE	Streptavidin; ligand; binding affinity; mutant; isolation;
KW	purification; recover; immobilise.
KM	
OS	Streptomycetes avidinii.
OS	Synthetic.
FH	Key Location/Qualifiers
FT	Misc-difference 68..71
FT	/label= VTAR68SSAV
FT	/note="Wild type ESAV is replaced by VTAR. Numbering is from the start of the mature protein"
PN	EP835934-A2.
PD	15-APR-1998.
PF	09-OCT-1997; 9TEP-0117504.
PR	10-OCT-1996; 96DE-1041876.
PA	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
PI	Skerra A, Voss S:
DR	N-PESDB; AAU34715.
PT	Streptavidin mutants with higher binding affinity for peptide ligands - have mutation in amino acid region 44-53, used to isolate, purify or determine fusion proteins including these ligands
PS	Disclosure; Page -: 21pp; German.
CC	This sequence represents a mutant streptavidin protein isolated from Streptomyces avidinii where the residues ESAV at position 44-47 of the mature wild type sequence are replaced by VTR. This sequence is used to produce mutants which are used in a method to assay the binding affinity of streptavidin mutants. These mutants have a mutation within the amino acid (aa) region 44-53 of the wild-type protein show a higher binding affinity than the wild-type for peptide ligands that include the sequence of formula Trip-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin mutants can be used to isolate, purify and determine proteins or to determine/recover substances that contain streptavidin-binding groups. Such compounds may also be used to immobilise fusions on microtitre plates, microbeads or sensor chips. NOTE: This sequence does not appear in the specification but has been constructed from the wild-type streptavidin sequence represented in AAU34714.
CC	
CX	
SC	Sequence 183 AA;
NY	Query Match 98.2%; Score 919; DB 19; Length 183; Best Local Similarity 98.4%; Pred. No. 3.1e-74; Matches 180; Conservative 1; Mismatches 2; Indels 0; Gaps 0
DG	
OY	1 MKRIIVAAIAVSITTSITASASADSKSQAQVAEAGTGTWNOLGSTIIVAAGD 60
DB	1 MKRIIVAAIAVSITTSITASASADS KSKAQVASAEAGTTWTWNOLGSTIIVAAGD 60
OY	61 GALTGRYESAVGAERSVVLTGRYDSAPTDDSGTLGWTVAKNNYNRNASHATTMSGY 120
DB	61 GALTGVTVARKNAERSRVLTGRYDSDPATDDSGTLGWTVAKNNYNRNASHATTMSGY 120
OY	121 VGCAEARINTOWLLTSTGEANAWKSLVGHDTFTRYVKPEASAIDAKKAQVNNGPLDA 180

```

Db      121  VCGAEARINQWLTSGTTEANMAKSLVGHDTFTKVPSPAASIDAKKGVANNGNFLDA 180
QY      181  VQQ 183
        |||
Db      181  VQQ 183

RESULT 14
AAMS59218
AAMS59218 standard; Protein, 183 AA.
XX
AC      AAMS59218;
XX
DT      27-AUG-1998 (first entry)
XX
DE      Streptavidin; ligand; binding affinity; mutant; isolation;
KW      purification; recover; immobilise.
XX
OS      Streptomyces avidinii.
XX
FH      Key location/Qualifiers
FT      Misc-difference 68..71
FT      /label= IGAR68ESAV
FT      /note= "Wild type ESAV is replaced by IGAR. Numbering
FT      is from the start of the mature protein"
XX
PN      EP835934-A2.
XX
PD      15-APR-1998.
XX
PF      09-OCT-1997; 97EP-0117504.
XX
PR      10-OCT-1996; 96DE-1041876.
XX
PA      (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
PI      Skerra A, Voss S;
XX
DR      WPI, 1998-218868/20.
XX      N-PDB; AAV34716.
XX
PT      Streptavidin mutants with higher binding affinity for peptide
PT      ligands - have mutation in amino acid region 44-53, used to isolate,
PT      purify or determine fusion proteins including these ligands
XX
PS      Disclosure; Page -: 21pp; German.
XX
CC      This sequence represents a mutant streptavidin protein isolated from
CC      Streptomyces avidinii where the residues ESAV at position 44-47 of
CC      the mature wild type sequence are replaced by IGAR. This sequence is
CC      used to produce mutants which are used in a method to assay the binding
CC      affinity of streptavidin mutants. These mutants have a mutation within
CC      the amino acid (aa) region 44-53 of the wild-type protein show a higher
CC      binding affinity than the wild-type for peptide ligands that include Y
CC      the sequence of formula  $Tyr-X-His-Pro-Gln-Phe-Y-Z$  where X = any aa; Y
CC      and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant
CC      streptavidin mutants can be used to isolate, purify and determine
CC      proteins or to determine/recover substances that contain
CC      streptavidin-binding groups. Such compounds may also be used to
CC      immobilise fusions on microtitre plates, microbeads or sensor chips.
CC      NOTE: This sequence does not appear in the specification but has
CC      been constructed from the wild-type streptavidin sequence represented
CC      in AAV34714.
XX
SQ      Sequence 183 AA;
XX
Query Match 98.0%; Score 917; DB 19; Length 183;
Best Local Similarity 98.4%; Pred. No. 4,7e-74;
Matches 180; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```


QY 1 MRKIIVAAIAVSLTIVSITASADPSKDSKAQVSAEAGITGTWYNOLGSTFIYTAGAD 60
 DB 1 MRKIIVAAIAVSLTIVSITASADPSKDSKAQVSAEAGITGTWYNOLGSTFIYTAGAD 60
 QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGGTALGWTVMKNNYRNAHSATTWSGOY 120
 DB 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGGTALGWTVMKNNYRNAHSATTWSGOY 120
 QY 121 VGGAEARINTQWLITSGTTEANAMKSTLVGHDTFTVKKPSAASIDAKKAGVNNGNPLA 180
 DB 121 VGGAEARINTQWLITSGTTEANAMKSTLVGHDTFTVKKPSAASIDAKKAGVNNGNPLA 180
 QY 181 VQOQ 183
 DB 181 VQOQ 183
 RESULT 15
 ID AAY84021 standard; protein; 186 AA.
 XX AAY84021;
 AC AAY84021;
 XX
 DT 03-JUL-2000 (first entry)
 XX
 DE Amino acid sequence of a fibrinonectin/streptavidin fusion.
 XX
 KM Streptavidin; monomer; biotin binding domain; functional domain;
 KM biotin; cell adhesion peptide; adaptor; streptavidin/biotin interaction;
 KM substrate; vascular device; prosthesis; fibrinonectin.
 XX
 OS Synthetic.
 OS Streptomyces avidinii.
 FH Key Location/Qualifiers
 FT Peptide 1..24
 FT /note= "propeptide"
 XX
 PN MO20001152-A1.
 XX
 PD 02-MAR-2000.
 XX
 PF 25-AUG-1999; 99MO-US19481.
 XX
 PR 25-AUG-1998; 98US-0097816.
 XX
 PA (UNIM) UNIV WASHINGTON.
 XX
 PI Stayton PS, McDevitt TC, Nelson KE;
 XX
 DR MPI, 2000-224689/19.
 XX
 PT Mutant streptavidin molecule, useful as an adaptor and a coating for
 PT substrates such as vascular devices or prostheses, comprises a biotin
 PT binding domain and a secondary functional domain -
 XX
 PS Example 4; Page 36-37; 43pp; English.
 CC The present sequence represents a fusion of fibrinonectin and
 CC streptavidin. It is a streptavidin molecule of the invention. The
 CC specification describes streptavidin molecules comprising a biotin
 CC binding domain and a secondary functional domain. The molecules are
 CC adaptors with inherent effector function. They can therefore, bind to
 CC biotin and also have another function, for e.g. binding to a cell
 CC through the secondary functional domain comprising a cell adhesion
 CC peptide. The streptavidin molecules are useful as adaptors to bring,
 CC via a streptavidin/biotin interaction, the secondary functional domain
 CC into proximity with a cell or molecule to be affected and as a coating
 CC for substrates such as vascular devices or prostheses. Therefore, any
 CC compound of interest, such as a nucleic acid, protein, peptide, organic
 CC compound, inorganic compound, polysaccharide or a combination, can be
 CC targeted, delivered or immobilized using them. The molecules, when
 CC comprising an antibody for its secondary functional domain, are also

CC useful in diagnostic applications for detecting analytes.

XX Sequence 186 AA;

Query Match 96.3%; Score 901.5; DB 21; Length 186;
 Best Local Similarity 96.2%; Pred. No. 1.2e-72;
 Matches 179; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

QY 1 MRKIIVAAIAVSLTIVSITASADPSKDSKAQVSAEAGITGTWYNOLGSTFIYTAGAD 60
 DB 1 MRKIIVAAIAVSLTIVSITASADPSKDSKAQVSAEAGITGTWYNOLGSTFIYTAGAD 60
 QY 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGGTALGWTVMKNNYRNAHSATTWS 117
 DB 61 GALTGTYESAVGNAESRYVLTGRYDSAPATDGGTALGWTVMKNNYRNAHSATTWS 120
 QY 118 GQYVGGAEARINTQWLITSGTTEANAMKSTLVGHDTFTVKKPSAASIDAKKAGVNNGNP 177
 DB 121 GQYVGGAEARINTQWLITSGTTEANAMKSTLVGHDTFTVKKPSAASIDAKKAGVNNGNP 180
 QY 178 LPAVQO 183
 DB 181 LPAVQO 186

Search completed: October 27, 2003, 10:44:00
 Job time : 44 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: October 27, 2003, 22:14:12 ; Search time 1931 Seconds
(without alignments)
2303.324 Million cell updates/sec

Title: US-09-589-870B-2
Perfect score: 936
Sequence: 1 MKRIYVAIAVSLTVSIRA.....IDAKKAGVNGNPLDAVQ 183

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_plus.model -DEV=xlh
-Q/cgn2_1/USPTO.epool/US0589870/runac_27102003_104446_16625/app_query.fasta_1.327
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=b1bs -START=1 -END=1 -MATRIX=b1bsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORX=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0589870.QCEN_1_1_2810@runac_27102003_104446_16625 -NCPU=6 -ICPU=3
-NO_MAP -LANG=QUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGIOLO
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST :
1: em_estba :
2: em_esthum :
3: em_estin :
4: em_estnu :
5: em_estrov :
6: em_estrpl :
7: em_estro :
8: em_hic :
9: gb_est1 :
10: gb_est2 :
11: gb_hic :
12: gb_est3 :
13: gb_est4 :
14: gb_est5 :
15: em_estfun :
16: em_estom :
17: em_gse_hum :
18: em_gse_inv :
19: em_gse_pin :
20: em_gse_vrt :
21: em_gse_fun :
22: em_gse_mam :
23: em_gse_mus :
24: em_gse_pro :
25: em_gse_rtd :
26: em_gse_png :
27: em_gse_vr1 :
28: gb_gse1 :

29: gb_gse2 :
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	165.5	17.7	435	12	BM080994
2	165.5	17.7	534	12	BM080994
3	162.5	17.4	598	12	BM183382
4	150.5	16.1	537	9	AL919330
5	150.5	16.0	406	12	BM531019
6	148	15.8	570	13	BU409120
7	148	15.8	578	13	BU426481
8	148	15.8	612	13	BU422332
9	147	15.7	538	12	BM488973
10	146.5	15.7	542	12	BI979837
11	146.5	15.7	608	13	BM263336
12	146	15.6	576	13	BU447569
13	145.5	15.5	622	13	BU293374
14	145	15.5	575	13	BU447766
15	145	15.5	582	13	BU408647
16	145	15.5	582	13	BM263637
17	145	15.5	777	13	BM385250
18	143.5	15.3	695	12	BM529407
19	143	15.3	601	12	BM508889
20	143	15.3	654	14	CB507800
21	141	15.1	568	14	CB507500
22	139.5	14.9	473	12	BM038085
23	132.5	14.2	588	13	BM081587
24	129.5	13.8	430	12	BM1708243
25	129	13.6	548	14	CB507054
26	127.5	13.6	592	12	BI672746
27	126	13.5	593	13	BU418594
28	124.5	13.3	809	13	BU386325
29	124.5	13.3	901	13	BU382470
30	124	13.2	771	13	BU383814
31	122.5	13.1	495	9	AM632946
32	122.5	13.1	505	13	BM264326
33	120.5	12.9	379	10	BM414084
34	115	12.3	614	13	BU410176
35	113.5	12.1	470	12	BM1709652
36	112	12.0	427	13	BM265638
37	110.5	11.8	342	14	CB490982
38	110.5	11.8	401	14	CB488233
39	109	11.6	453	13	BM278781
40	108.5	11.6	346	14	BM486748
41	108.5	11.6	352	14	BM487730
42	108.5	11.6	367	14	CB491403
43	108.5	11.6	376	14	CB486669
44	108.5	11.6	381	14	CB490090
45	108.5	11.6	384	14	CB487378

ALIGNMENTS

RESULT 1
LOCUS BM080994
DEFINITION BM080994 435 bp mRNA linear EST 14-NOV-2001
IMAGE:5159603 5' similar to SW:AVID_CHICK P02701 AVIDIN PRECURSOR.
[1] : mRNA sequence.
ACCESSION BM080994
KEYWORDS EST.
SOURCE Danio rerio (zebrafish)
ORGANISM Danio rerio
Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes

REFERENCE
AUTHORS
1 (bases 1 to 435)
Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE
JOURNAL
COMMENT
Washu Zebrafish EST Project 1998
Unpublished
Other ESTs: ft78g06.x1
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.wustl.edu
The library was constructed by Dr. Z. Gong. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis. Please contact Zhiyuan Gong for further information on this library (National University of Singapore, Department of Biological Sciences, Lower Kent Ridge Road, Singapore 119260).
High quality sequence stop: 393.
Location/Qualifiers
1..435
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:5159603"
/sex="female"
/dev_stage="4-5 month"
/lab_host="MD10B (phage-resistant)"
/clone_lib="Gong zebrafish ovary"
/note="Organ: ovary (pooled); Vector: pBluescript SK-1; Site_1: XhoI; Site_2: EcoRI; Poly A+ RNA was isolated from the ovaries of 2 female adult zebrafish (4-5 month old). cDNAs were made using oligo-dT primers and inserted into lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo mass-excised to pBluescript SK- following the Washington University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Zhiyuan Gong for further information on this library (National University of Singapore, Department of Biological Sciences, Lower Kent Ridge Road, Singapore 119260)."

BASE COUNT
99 a 100 c 131 g 105 t

ORIGIN
Alignment Scores:
Pred. No.: 7.8e-08 Length: 435
Score: 165.50 Matches: 54
Percent Similarity: 48.80% Conservative: 27
Best Local Similarity: 32.53% Mismatches: 50
Query Match: 17.68% Indels: 35
DB: 12 Gaps: 10

US-09-589-870b-2 (1-183) x BM080994 (1-435)

QY 18 IIEThRALaSerAlaSerlAspProSerlyAspSerlyAlaGlnValSerAla 37
Db 3 ATCTCAAAAGCAGCGCCGCAACGAGTCTCAACGCCCAATGCAAGGAGCTCC-- 59
QY 38 GUAAlaGlylIeThrGlyThrTrpYAsGlnIleuGlySerTherPheIleValThrala 57
Db 60 ---TGTAAATGTCACCGGTCTTTGGCCCATGAGCTTGCTTACGCTGGGGTGAAGGCG 116
QY 58 ---GlyAlaAspGlyAlaLeuThrGlyThrTrpYAsGlnIleuGlySerTherPheIleValThrala 76
Db 117 GAAGGCTCGAG-----GTCAAGAGGTGTTTACCAAGACCGCGGTT-----GAGAGT 161
QY 77 AgtGlyValLeuThrGlyArgTrpAspSerAla-----ProAlaThrAspGly 92
Db 162 ACGCAGGAGCGCGCAGGTATCATCGCTCCGCCGATCATCGGAGTGTCTAGTACGGG 221

QY 93 SerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnTyArgAsnAlaHisSer 112
Db 222 ACTCAGCCTACCGTCTCTTCTCCGTGCTATGGAGAAAGT-----TCG 266
QY 113 AlaThrThrTrpSerGlyGlnTrpVal-----GlyGlyAlaGlnAlaArgIleAsn 129
Db 267 TGCCTGCTGCTGGGTGGTGGCAGTCTTATTATTAGATGACGGGACACAGTGC--CTGAAA 323
QY 130 ThrGlnTrpLeuLeuThrSer---GlyThrThrGlnAlaAsnAlaTrpLysSerThrLeu 148
Db 324 ACCTTTGGATGTTGGCAGGCTTGGCAGACAAATTGGCAATGCTCTGGGACACCCGA 383
QY 149 ValGlyHisAspTrpPheThrLysValLysProSerAlaAlaSerIleAspAlaLys 168
Db 384 ATGGAGAAAGATATTTTTTC----- 404
QY 169 LysAlaGlyValaAsnAsn 174
Db 405 AAGACTGAGATTCCAAT 422

RESULT 2
LOCUS
DEFINITION
B1980006/c 534 bp mRNA linear EST 24-OCT-2001
ft78g06.x1 Gong zebrafish ovary Danio rerio cDNA clone
IMAGE:5159603 3' similar to SW:AVID_CHICK P02701 AVIDIN PRECURSOR.
[1] :, mRNA sequence.
ACCESSION
B1980006
VERSION
B1980006.1 GI:16367292
KEYWORDS
EST.
SOURCE
Danio rerio (zebrafish)
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Osteichthys; Cypriniformes
; Cyprinidae; Danio.
1 (bases 1 to 534)
Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE
JOURNAL
COMMENT
Washu Zebrafish EST Project 1998
Unpublished
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.wustl.edu
The library was constructed by Dr. Z. Gong. DNA Sequencing by: Washington University Genome Sequencing Center St. Louis. Please contact Zhiyuan Gong for further information on this library (National University of Singapore, Department of Biological Sciences, Lower Kent Ridge Road, Singapore 119260).
Seq primer: 17 from Gidco
High quality sequence stop: 421.
Location/Qualifiers
1..534
/organism="Danio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:5159603"
/sex="female"
/dev_stage="4-5 month"
/lab_host="MD10B (phage-resistant)"
/clone_lib="Gong zebrafish ovary"
/note="Organ: ovary (pooled); Vector: pBluescript SK-1; Site_1: XhoI; Site_2: EcoRI; Poly A+ RNA was isolated from the ovaries of 2 female adult zebrafish (4-5 month old). cDNAs were made using oligo-dT primers and inserted into lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo mass-excised to pBluescript SK- following the Washington

University protocol
(http://genome.wustl.edu/esf/lambda_protocol.shtml).

Please contact Zhiyuan Gong for further information on this library (National University of Singapore, Department of Biological Sciences, Lower Kent Ridge Road, Singapore 119260).

BASE COUNT 142 a 154 c 115 g 123 t
ALIGNMENT SCORES:
Pred. No.: 1.04e-07 Length: 534
Score: 165.50 Matches: 54
Percent Similarity: 48.80% Conservative: 27
Best Local Similarity: 32.53% Mismatches: 50
Query Match: 17.68% Indels: 35
DB: 12 Gaps: 10

US-09-589-870B-2 (1-183) x B1980006 (1-534)

QY 16 TleThAlaSerAlaSerAlaAproSerLyAapSerLyAaGlnValSerAlaA 37
DB 531 ATCTCAAAAGCAGCGTCGCCAAGAGCTCTCAACGCCCAATGCAAGGTGAGCTCC--- 475
QY 38 GUAAGAGlyleThrGlyThrTPYrAaGlnLeuGlySerThrPheIleValThraA 57
DB 474 ---TGTAATGTCAACCGGTGTTGGCCCAATGAGCTTGCTCTACGCTCGCGGTAAGCGC 418
QY 58 ---GlyAlaAapGlyAlaLeuThrGlyThrTyGluSerAlaValGlnAaGlnSer 76
DB 417 GAAGGCTCGGAG---GTCAAGAGGTGTTTACACAGCCCGGCTT-----GAGAGT 373
QY 77 ArgTyValLeuThrGlyThrTyAapSerAla-----ProAlaThraPgly 92
DB 372 ACCCAGGAGGCCAGGATCATCCCTCGCCGATCATCGGATGCTGAGTACGGG 313
QY 93 SerGlyThraLeuGlyThrThraValAlaThrLyAaAaAaThraAaAaSer 112
DB 312 ACTCAGCCTACCGCTCTTCTCCGCTGCTATGGAGAAAGT-----TGC 268
QY 113 AlaThrThrTySerGlyGlnTyVal-----GlyGlyAlaGlnAlaAaGln 129
DB 267 TGCCTGCTGCGGTGCTGCTGCTTATTTAGATACAGCGGACAGAGT---CTGAA 211
QY 130 ThGlnThrLeuThrSer---GlyThrThraAlaAaAaThraPlySerThra 148
DB 210 ACTTGTGATGTGGCAGATGTGGCAGACATTTGGCAAGTCCGCGGAGACACAGA 151
QY 149 ValGlyAaAaThrPheThrValLyPProSerAlaAaSerIleAaAaAaLy 168
DB 150 ATGGAGAAAGATATTTTTC----- 130
QY 169 LyAaGlyValAaAaAa 174
DB 129 AAGACTGAGATTCAAT 112

RESULT 3
BM183382 598 bp mRNA linear EST 11-DEC-2001
LOCUS fV63a05.y1 Sugano SJD adult male Dantio rerio cDNA clone
DEFINITION IMAGE:541336.5' similar to SW:AVID_CHICK P02701 AVIDIN PRECURSOR.
[1] : mRNA sequence.
ACCESSION BM183382
VERSION BM183382.1 GI:17514340
KEYWORDS EST.
SOURCE Dantio rerio (zebrafish)
ORGANISM Dantio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Dantio.
1 (bases 1 to 598)
REFERENCE
AUTHORS Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Peterson, B.,

TITLE
JOURNAL
COMMENT
Unpublished
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafi@wustl.edu

Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritzer, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.
WashU Zebrafish EST Project 1998
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA Sequencing by Washington University Genome Sequencing Center Clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, send email to: info@image.llnl.gov
Seq primer: T3 ET from Amersham
High quality sequence stop: 527.
Location/Qualifiers

FEATURES
source

1..598
/organism="Dantio rerio"
/mol_type="mRNA"
/db_xref="taxon:7955"
/clone="IMAGE:541336"
/sex="male"
/tissue_type="whole body"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/clone_id="Sugano SJD adult male"
/note="Vector: pME18S-FL3; Site 1: DraIII (CACCATGTG); Site 2: DraIII (CACTGTGTG); 1st strand cDNA was primed with an oligo(dT) primer (ATGCGCCCTTTTCTTTTCTTTT); double-stranded cDNA was ligated to a DraIII adaptor (TTTGGGCTACTG), digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed and donated by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTTAAAGCTGCG and 3' end primer CGACCTGACGCTCGAGCA."

BASE COUNT 140 a 122 c 167 g 169 t

ALIGNMENT SCORES:

Pred. No.: 2.54e-07 Length: 598
Score: 162.50 Matches: 55
Percent Similarity: 47.95% Conservative: 27
Best Local Similarity: 32.16% Mismatches: 54
Query Match: 17.36% Indels: 35
DB: 12 Gaps: 10

US-09-589-870B-2 (1-183) x BM183382 (1-598)

QY 13 LeuThrThraValSerIleThraLeuThraAaAaAaAaAaAaAaAaAaAaAa 32
DB 37 CTCTGAGCTTGTGACCTCAAAAGCTCGTCCGAAAGCTCTCAACGCCAATGCA 96
QY 33 GlnValSerAlaAaGlnAlaGlyleThrGlyThrTPYrAaGlnLeuGlySerThr 52
DB 97 AAGGTGAGCTCC-----TGTAATGTCAACCGGTGTTGGCCCAATGAGCTTGCTTACG 150
QY 53 PheIleValThraAa---GlyAlaAaAaAaAaAaAaAaAaAaAaAaAaAaAaAa 71
DB 151 CTGCGGTGAAGCGCGAAGGCTCGAG-----GTCAAGAGGTGTTTACACAGCCGGTT 204
QY 72 GlyAaAaAaAaGlnSerAlaGlyThrValLeuThrGlyAaGlyThrTyGluSerAlaVal 71
DB 205 -----GAGAGTACCGCGCGAGCCAGGATCATACCGCTCGCGCATCATCGCG 87
QY 88 ProAlaThraAa 255
DB 256 ATGCTCAGTACGAGGAGCTACAGCTTCTTCTCGTCTATGGAGAAAGT--- 312

QY 108 ArgAsnAlaHisSerAlaThrThrPserGlyIntyVal-----GlyGlyAla 124
 Db 313 -----TCGTCCTCAGCTGGCTGGTGGTCACTGTTATTATTATGATACGGGGGA 360
 QY 125 GluAlaArgIleAsnThrGlnTrpLeuLeuThrSer---GlyThrGluAlaAsnAla 143
 Db 361 CAAGTG---CTGAAACCTTCTGATGTTGCCAGCTGTTGCAGACAATTGGCAAGTCC 417
 QY 144 TrpLysSerThrLeuValGlyHisAspThrPheThrLysValLysProSerAlaIleSer 163
 Db 418 TGGGAGAGACCAAGATGGAGAGATATTATTTTTC----- 453
 QY 164 IleAspAlaAlaLysLysAlaGlyValaAsn 174
 Db 454 -----AAGACTGAGTTCAAT 471
 RESULT 4
 AL919330/c 537 bp mRNA linear EST 18-SEP-2002
 LOCUS AL919330 PUR-Z1+22 Danio rerio cDNA clone 061-G04-2, mRNA sequence.
 DEFINITION AL919330
 ACCESSION AL919330
 VERSION AL919330.1 GI:23184628
 KEYWORDS EST
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
 ; Cyprinidae; Danio.
 1 (bases 1 to 537)
 Lee, S., Ruan, H., Ma, W., Xu, M., Lo, J., He, Y., Liu, F., Eun, A., Wen, Z.
 and Peng, U.
 15,000 unique zebrafish EST clusters from two cDNA libraries
 Contact: Peng U
 Unpublished
 TITLE Lab of Functional Genomics
 JOURNAL Institute of Molecular and Cell Biology
 COMMENT 30 Medical Drive, Singapore, 117609, Singapore
 Email: pengj@imcb.a-star.edu.sg
 Clone requests: pengj@imcb.a-star.edu.sg.
 Location/Qualifiers
 FEATURES
 source
 1..537
 /organism="Danio rerio"
 /mol_type="mRNA"
 /strain="local wildtype"
 /db_xref="taxon:7955"
 /clone="061-G04-2"
 /tissue_type="whole embryo or fish"
 /dev_stage="mixed stages"
 /clone_1lb="PUR-Z1+22"
 BASE COUNT 145 a 146 c 107 g 139 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 414e-06 Length: 537
 Score: 150.50 Matches: 49
 Percent Similarity: 48.68% Conservative: 25
 Best Local Similarity: 32.24% Mismatches: 43
 Query Match: 16.08% Indels: 35
 DB: 9 Gaps: 10
 US-09-589-870b-2 (1-183) x AL919330 (1-537)
 QY 32 AlaGlnValSerAlaAlaGluAlaGlyIleThrGlyThrTrpTyraSngInLeuGlySer 51
 Db 527 GCAAGGAGAGCTCC-----TGTATGTCACCGGTGTTGGCACATGAGCTTGGCTCT 474
 QY 52 ThrPheIleValThrAla---GlyAlaAspGlyAlaLeuThrGlyThrTyGluSerAla 70
 Db 473 ACGCTGCGGCTGACAGCGAGAGCTCGAG-----GTCAGAGCTGTTTACACAGCCGCG 420
 QY 71 ValGlnAsnAlaGluSerArgTyrrValLeuThrGlyArgTyraSpsrAla----- 87

Db 419 GTT-----GAGAGTACCCAGGAGCCGAGCTATCACCGCTCCGCCGATCATC 369
 QY 88 ---ProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsn 106
 Db 368 GGGATGTCAGTACGAGGACTACGACCTACCTCTTCTTCGTCGATAGGAGAAAGT 309
 QY 107 TyrArgAsnAlaHisSerAlaThrThrTrpSerGlyIntyVal-----GlyGly 123
 Db 308 -----TCGTCCTCAGCTGGCTGGTGGTCACTGTTATTATTATGATACGGG 264
 QY 124 AlaGluAlaArgIleAsnThrGlnTrpLeuLeuThrSer---GlyThrGluAlaAsn 142
 Db 263 GCACAGTA---CTGAAACCTTCTGATGTTGCCAGCTGTTGCAGACAATTGGCAAGT 207
 QY 143 AlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValLysProSerAlaAla 162
 Db 206 CCTGGGAGACACACAGATGGAGAGATTTTTC----- 168
 QY 163 SerIleAspAlaAlaLysLysAlaGlyValaAsn 174
 Db 167 -----AAGACTGAGTTCAAT 150
 RESULT 5
 BM531019 406 bp mRNA linear EST 19-FEB-2002
 LOCUS BM531019
 DEFINITION fW9f04.Y1 Gong zebrafish ovary Danio rerio cDNA clone
 IMAGE:5618455, similar to SW:AVID_CHICK P02701 AVIDIN PRECURSOR.
 [1] : mRNA sequence.
 ACCESSION BM531019
 VERSION BM531019.1 GI:187939320
 KEYWORDS EST
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
 ; Cyprinidae; Danio.
 1 (bases 1 to 406)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Maria, M., Eddy
 S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
 K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
 Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R., Ritter, E.,
 Korn, S., Shin, T., Jackson, Y., Cardenas, M., McCall, R., Waterston, R.
 and Wilson, R.
 Washu Zebrafish EST Project 1998
 Unpublished
 TITLE Washu Zebrafish EST Project 1998
 JOURNAL
 COMMENT Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrfish@wustl.edu
 The library was constructed by Dr. Z. Gong. DNA sequencing by:
 Washington University Genome Sequencing Center St. Louis. Please
 contact Zhiyuan Gong for further information on this library
 (National University of Singapore, Department of Biological
 Sciences, Lower Kent Ridge Road, Singapore 119260).
 Seq primer: T3 BT from Amersham
 High quality sequence stop: 395.
 Location/Qualifiers
 FEATURES
 source
 1..406
 /organism="Danio rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="IMAGE:5618455"
 /sex="female"
 /dev_stage="4-5 month"
 /lab_host="DH10B (phage-resistant)"
 /clone_1lb="Gong zebrafish ovary"
 /note="Organ: ovary (pooled); Vector: pBluescript SK-;
 Site 1: XhoI; Site 2: EcoRI; Poly A+ RNA was isolated from
 the ovaries of 2 female adult zebrafish (4-5 month old).
 cDNAs were made using oligo-dT primers and inserted into
 lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo

Db 502 CGCAAGCCAGCAACATGCC 522

RESULT 7
BU426481 578 bp mRNA linear EST 29-NOV-2002
LOCUS 603958787f1 CSEQRBNO9 Gallus gallus cDNA clone CHEST930n22 5', mRNA
DEFINITION sequence.

ACCESSION BU426481
BU426481.1 GI:25919157
VERSION EST.
KEYWORDS Gallus gallus (chicken)
SOURCE Gallus gallus
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 578)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.D.,
A comprehensive collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers

1. 578
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer and broiler"
/db_xref="taxon:9031"
/clone="CHEST930n22"
/sex="Male and female"
/tissue_type="Chondrocytes isolated from growth plate cartilage"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQRBNO9"
/note="Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., FNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

BASE COUNT 152 a 171 c 145 g 110 t
ORIGIN

Alignment Scores:
Pred. No.: 8.48e-06 Length: 578
Score: 148.00 Matches: 48
Percent Similarity: 46.26% Conservative: 20
Best Local Similarity: 32.65% Mismatches: 59
Query Match: 15.81% Indels: 20
DB: 13 Gaps: 8

US-09-589-870b-2 (1-183) x BU426481 (1-578)

QY 34 valserAlaIaagUaIaGlyIleThrGlyThrTyrPyrAsnGlnleuGlySerThrPhe 53
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 97 CTCTTCGCAGAAAGTCTGCTGACTGAGAAATGACCAACATCTGGGCTCAACATG 156

QY 54 lIleValThrAla---GlyAlaAspGlyAlaIleuThrGlyThrTyrGlyuSerAlaValGly 72
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 157 ACCATCGGGGTGTGAACCAAGGTGAATTCAACAGACCTTACACAGCGGTA--- 213
QY 73 AsnAlaGluSerArgTyrValIleuThrGlyArgTyrAspSerAlaProAlaThrAspGly 92
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 214 ACAGCCACATCAATGATGATCAAGTCAAGTCAACCTGATGGAGACAAACCATCAAC 273
QY 93 SerGlyThr-----AlaIleuGlyTyrThrValAlaIleuThrPyrAsnAsnThrArgAsnAla 110
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 274 AAGAGACCCAGCCACCTTGGCTCACTGCTCAATTGGAAG-----TTTTC 321
QY 111 HisSerAlaThrThrTyrSerGlyGlnTyrVal-----GlyGlyAlaGluAlaArg 127
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 322 GAGTCCACCACTGCTTCAAGCGGCGAGTCTTCATAGACAGAAACGGAAAGAGGTC-- 378
QY 128 lIleAsnThrGlnTyrPleuLeuThrSerGlyThrThrGlu---AlaAsnAlaThrPyrSer 146
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 379 CTGAAGACCATGTGGCTGCTCGGTCAAGTGTATGATGATGGTGAAGACT 438
QY 147 ThrIleuValGlyHisAspThrPheThrIleValys-----Pro 159
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 439 ACCAGGCTGGCATCAACATCTTCACTGCTCGGCACACAGAAAGACTGATGGCC 498
QY 160 SerAlaAlaSerIleAspAla 166
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
Db 499 CGCAAGCCAGCAACATGCC 519

RESULT 8
BU422332 612 bp mRNA linear EST 29-NOV-2002
LOCUS 603232324f1 CSEQRBNO9 Gallus gallus cDNA clone CHEST229p5 5', mRNA
DEFINITION sequence.

ACCESSION BU422332
BU422332.1 GI:25915008
VERSION EST.
KEYWORDS Gallus gallus (chicken)
SOURCE Gallus gallus
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 612)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.D.,
A comprehensive collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers

1. 612
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer and broiler"
/db_xref="taxon:9031"
/clone="CHEST229p5"
/sex="Male and female"
/tissue_type="Chondrocytes isolated from growth plate cartilage"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQRBNO9"
/note="Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated

using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

```
BASE COUNT      171 a      176 c      150 g      115 t
ORIGIN
Alignment Scores:
Pred. No.:      9.19e-06      Length:      612
Score:          148.00      Matches:      48
Percent Similarity: 46.26%      Conservative: 20
Best Local Similarity: 32.65%      Mismatches: 59
Query Match:    15.81%      Indels:      20
DB:             13      Gaps:      8

US-09-589-870b-2 (1-183) x BU422332 (1-612)
QY      34 ValSerAlaAlaGluAlaGlyIleThrGlyThrTyraSngInleuGlySerThrPhe 53
      99 CTCTTCGCCAGAAAGCTCGCTGACGGAATGACCAACATCTGGGCTCCACATG 158
      54 ILevalThrAla---GlyAlaAepGlyAlaLeuThrGlyThrTyrguSerAlaValGly 72
      159 ACCATCGGGCGCTGAAACAGCAAGGTAATTCACAGACCACTACACAGCCGTA--- 215
QY      73 AsnAlaGluSerArgTyraValLeuThrGlyArgTyraSperAlaProAlaThrAepGly 92
      216 ACAGCCACATCAATGATCAAGAGTCAACAGCTGAGTGGAGACCAAAACACCATCAAC 275
QY      93 SerGlyThr-----AlaLeuGlyTPrpThrValAlaIleThrLysAbnAenTyraTgaAsnAla 110
      276 AAGAGACCCAGCCACCTTGGCTTCACTGCTCAATTGGAG-----TTTTCa 323
QY      111 HisSerAlaThrThrTPrpSerGlyGlnTyraVal-----GlyGlyAlaGluAlaArg 127
      324 GAGTCCACACACTGCTTCAACGGGCGAGTGTCTTACAGACAGAAAGGAGAGGTC--- 380
QY      128 IleAenThrGlnTrpLeuLeuThrSerGlyThrThrGlu---AlaAsnAlaTrpLysSer 146
      381 CTGAAGCCATGTGGCTGCTGCGCTCAAGTGAATGACATGCTGATGCTGAAAGCT 440
QY      147 ThrLeuValGlyHisAepThrPheThrLysValLys-----Pro 159
      441 ACCAGGTCGGCATCAACATCTTCACTCGCTCGGCACACAGAAAGAGTGGATGGCC 500
DB      160 SerAlaAlaSerIleAspAla 166
      501 CGCAAGCCAGCAACATGCC 521

RESULT 9
LOCUS      BM489973      538 bp      mRNA      linear      EST 07-FEB-2002
DEFINITION      pBp2n.pK002.a9 Normalized Chicken Pituitary/Hypothalamus/Pineal
      library (pBp2n) Gallus gallus cDNA clone pBp2n.pK002.a9 5' similar
      to emb|C6C34569.1 (A0111647) avadin [Gallus gallus], mRNA sequence.
ACCESSION      BM489973
VERSION      BM489973.1      GI:18610904
KEYWORDS      EST.
SOURCE      Gallus gallus (chicken)
ORGANISM      Gallus gallus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE      1 (bases 1 to 538)
AUTHORS      Porter,T.E. and Cogburn,L.A.
TITLE      ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
```

library, USDA/IFAPS Animal Genome Project
Unpublished
Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu www.chickest.udel.edu.
Location/Qualifiers
1.538
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Commercial broiler chickens"
/db_xref="taxon:9031"
/clone="pBp2n.pK002.a9"
/sex="Male and Female"
/tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
/dev_stage="Embryonic (d12,d14,d19) ; post-hatch (1,3,5,7,9 weeks)"
/lab_host="E. coli EMD10B"
/clone_lib="Normalized Chicken
Pituitary/Hypothalamus/Pineal Library (pBp2n)"
/note="Vector: PCWVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue at different
ages. Single pass sequencing from 5'-end"

```
BASE COUNT      140 a      157 c      137 g      104 t
ORIGIN
Alignment Scores:
Pred. No.:      9.8e-06      Length:      538
Score:          147.00      Matches:      48
Percent Similarity: 46.26%      Conservative: 20
Best Local Similarity: 32.65%      Mismatches: 59
Query Match:    15.71%      Indels:      20
DB:             12      Gaps:      8

US-09-589-870b-2 (1-183) x BM489973 (1-538)
QY      34 ValSerAlaAlaGluAlaGlyIleThrGlyThrTyraSngInleuGlySerThrPhe 53
      63 CTCTTCGCCAGAAAGCTCGCTGACGGAATGAGACAGCACTGGGCTCCACATG 122
      54 ILevalThrAla---GlyAlaAepGlyAlaLeuThrGlyThrTyrguSerAlaValGly 72
      123 ACCATCGGGCGCTGAAACAGCAAGGTAATTCACAGGACCTACCAAGCCGTA--- 179
QY      73 AsnAlaGluSerArgTyraValLeuThrGlyArgTyraSperAlaProAlaThrAepGly 92
      180 ACAGCCACATCAATGATCAAGAGTCAACAGCTGAGTGGAGACCAAAACACCATCAAC 239
QY      93 SerGlyThr-----AlaLeuGlyTPrpThrValAlaIleThrLysAbnAenTyraTgaAsnAla 110
      240 AAGAGACCCAGCCACCTTGGCTTCACTGCTCAATTGGAG-----TTTTCa 287
QY      111 HisSerAlaThrThrTPrpSerGlyGlnTyraVal-----GlyGlyAlaGluAlaArg 127
      288 GAGTCCACACACTGCTTCAACGGGCGAGTGTCTTACAGACAGAAAGGAGGTC--- 344
QY      128 IleAenThrGlnTrpLeuLeuThrSerGlyThrThrGlu---AlaAsnAlaTrpLysSer 146
      345 CTGAAGACCATGTGGCTGCTGCGCTCAAGTGAATGACATGCTGATGCTGAAAGCT 404
QY      147 ThrLeuValGlyHisAepThrPheThrLysValLys-----Pro 159
      405 ACCAGGTCGGCATCAACATCTTCACTCGCTCGGCACACAGAAAGAGTGGATGGCC 464
DB      160 SerAlaAlaSerIleAspAla 166
      465 CGCAAGCCAGCAACATGCC 485

RESULT 10
LOCUS      B1979837      542 bp      mRNA      linear      EST 24-OCT-2001
```

DEFINITION ft76d11.y1 Gong zebrafish ovary Danio rerio cDNA clone
IMAGE:5159109.5, similar to SW:AVID_CHICK P02701 AVIDIN PRECURSOR.
[1], mRNA sequence.

ACCESSION BI979837.1 GI:16366965

VERSION BI979837.1

KEYWORDS EST.

SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

AUTHORS Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucba, T., Martin, J., Beck, C., Wyllie, T., Underwood, K., Septoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

TITLE WashU Zebrafish EST Project 1998

COMMENT Unpublished

CONTACT: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu

The library was constructed by Dr. Z. Gong. DNA sequencing by:
Washington University Genome Sequencing Center St. Louis. Please
contact Zhiyuan Gong for further information on this library
(National University of Singapore, Department of Biological
Sciences, Lower Kent Ridge Road, Singapore 119260).
Seq primer: T3 ET from Amersham

High quality sequence stop: 395.

Location/Qualifiers

1..342

/organism="Danio rerio"

/mol_type="mRNA"

/db_xref="taxon:7955"

/clone="IMAGE:5159109"

/sex="female"

/dev_stage="4-5 month"

/lab_host="PH10B (phage-resistant)"

/clone_lib="Gong zebrafish ovary"

/note="Organ: ovary (pooled); Vector: pBluescript SK-
Site 1: XhoI; Site 2: EcoRI; Poly A+ RNA was isolated from
the ovaries of 2 female adult zebrafish (4-5 month old).
cDNAs were made using oligo-dT primers and inserted into
lambda ZAP II vector (Stratagene) by Dr. Z. Gong, in vivo
mass-excised to pBluescript SK- following the Washington
University protocol
(http://genome.wustl.edu/est/lambda_protocol.shtml).
Please contact Zhiyuan Gong for further information on
this library (National University of Singapore,
Department of Biological Sciences, Lower Kent Ridge Road,
Singapore 119260)."

BASE COUNT 142 a 105 c 148 g 147 t

ORIGIN

Alignment Scores: 1.12e-05 Length: 542

Pred. No.: 146.50 Matches: 49

Score: 47.68% Conservative: 23

Percent Similarity: 47.68% Mismatches: 44

Best Local Similarity: 32.45% Indels: 35

Query Match: 15.65% Gaps: 10

DB: 12

US-09-589-870b-2 (1-183) x BI979837 (1-542)

QY 33 GlnValSerAlaIaGuaIaGlyIleThrGlyThrTrpTyraSngInleugySerTrn 52

DB 5 CAGGTAGCTCC-----TGTATGTACACCGGTTTGGCGCAATAGAGCTTGCTCAG 58

QY 53 PheIleValThrAla---GlyAlaSepGlyAlaLeuThrGlyThrTygIuSerAlaVal 71

DB 59 CTCGGCGTGAAGCGGAAAGCTCGAG-----CTCAGAGCTGTTTACAGACCGCGTT 112

QY 72 GlyAsnAlaGluSerArgTyValLeuThrGlyArgTyAspSerAla----- 87

DB 113 -----GAGAGTACCGACGAGCCGACGATATCCGCTCGCCGATCATCGCG 163

QY 88 ProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThrValAlaTrpTyAsnAsnTy 107

DB 164 ATGCTACGTACGCGGACCTACCGCTCTCTTCTCGGTATGGAGAAAGT--- 220

QY 108 ArgAsnAlaHisSerAlaThrThrTySerGlyGlnTyVal-----GlyGlyAla 124

DB 221 -----TCGTGCTCGCTTGCTGGTGTGTCAGTCTTATTTAGATGACGGGCA 268

QY 125 GluAlaArgIleAsnThrGlnTrpLeuThrSer---GlyThrThrGluAlaAsnAla 143

DB 269 CAGGTG---CTGAATCCTTCTGTGATGTTGGCGAGTGTGACAGCAATTGGCAAGTCC 325

QY 144 TrpTySerThrLeuValGlyHisAspThrPheThrTyValTyProSerAlaIaSer 163

DB 326 TCGGAGACACACAGATGGAGAGATATTTTTC----- 361

QY 164 IleAspAlaAlaLysLysAlaGlyValAsnAsn 174

DB 362 -----AGACTGAGATTCAAT 379

RESULT 11

LOCUS BX263336 608 bp mRNA linear EST 27-FEB-2003

DEFINITION BX263336 AGENAE Gallus gallus multi-tissues normalized library (gcag) Gallus gallus cDNA clone gcag0016c.g.24 5prim, mRNA sequence.

ACCESSION BX263336

VERSION BX263336.1 GI:28585934

KEYWORDS EST

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianidae; Gallus.

AUTHORS Herault, F., Le Menth-Metzinger, V., Desert, C., Retout, E., Piumi, F., Klopp, C. and Douaire, M.

TITLE Construction and primary characterization of chicken normalized multi-tissue cDNA libraries

JOURNAL Unpublished

COMMENT Contact: Douaire M

INRA, UMR INRA-ENSAR Genetique Animale
65, rue de Saint-Brieuc, RENNES cedex, 35042, FRANCE
Tel: +33 (0) 2.23.48.54.63
Fax: +33 (0) 2.23.48.54.70
Email: Madeleine.Douaire@roazon.inra.fr

Sequence cleaned of vector, adaptor and repetitions. Contact us at signasupport@jouy.inra.fr to obtain the chromatogram of this sequence.

Place: 0016 row: g column: 24

Seq primer: M13.

Location/Qualifiers

1..608

/organism="Gallus gallus"

/mol_type="mRNA"

/db_xref="taxon:9031"

/clone="gcag0016c.g.24"

/tissue_type="multi-tissues"

/tissue="from embryos to adults"

/dev_stage="from embryos to adults"

/lab_host="DH10B"

/clone_lib="AGENAE Gallus gallus multi-tissues normalized library (gcag)"

/note="Vector: pTTT3D-pac; tissues: brain, embryos, kidney, adipose tissue, granulosa, utero-vaginal gland, oviduct, small follicle, ovary, hypothalamus, pituitary gland,

lleon, jejenum, caecum, duodenum, spleen, fabricius gland
bone marrow, thymus, hematopoietic progenitor cells.
Clone distribution : AGENAE Resource centre, Francois
PIMI, Francois PIMI, intra.fr, INRA, CEA Radiobiologie et
Etude du genome (LREG), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, FRANCE"

BASE COUNT 111 a 204 c 179 g 114 t

ALIGNMENT SCORES:

Pred. No.: 1,31e-05 Length: 608
Score: 146.50 Matches: 50
Percent Similarity: 42.46% Conservative: 26
Best Local Similarity: 27.93% Mismatches: 62
Query Match: 15.65% Indels: 41
DB: 13 Gaps: 9

US-09-589-870B-2 (1-183) x BX263336 (1-608)

QY 34 ValSerAlaAlaGluAlaGlyIleThrGlyThrTrpTyraGlnLeuGlySerThrPhe 53
DB 80 GTGGCTCCGGGAGAGTCGAGAGGCTGTGGAGAGACGAGCTGGGCTCAACATG 139
QY 54 IleValThrAla---GlyAlaSerGlyAlaLeuThrGlyThrTyrgluserAlaValGly 72
DB 140 ACCATCTGACCTGGATGCTGGAGACCTTCCGGCTTACACAGACGGGTG--- 196
QY 73 AsnAlaGluSerArgTyraValLeuThrGlyArgTyraSerAla-----ProAla 89
DB 197 ACAGCCACCAACAGAGATCTGTGTCAACCCCTGAAGAGGCGCAGACCCCGCGC 256
QY 90 ThrAspGlySerGlyThrAlaLeuGlyThrTrpValAlaThrTyraSerAlaGln 109
DB 257 ACCAAGGGGAGAGACCCACCTTGGCTTCTACTGTGAGTGG-----CAGTTT 304
QY 110 AlaHisSerAlaThrThrTrpSerGlyGln---TyraValGly-----GlyAlaGluAla 126
DB 305 GCAGACTCCACCACTGCTTGTGGACAGAGCTGTGTGACCGCGTGGAGAGAG--- 361
QY 127 ArgIleAsnThrGlnTrpLeuLeuThrSerGlyThr---ThrGluAlaAsnAlaTrpLys 145
DB 362 ATCTGTGAGATGCGATGTTGCTGCGGAGAGAGTTCCTCCGCAAGGACACCTGGAAG 421
QY 146 SerThrLeuValGlyHisAspThrPheThrLysValLys----- 158
DB 422 GCCACGAGGTGGACCAATGTTTCAACCGGTGCAAGTATGGGAGACACTGCTGGC 481
QY 158 ----- 158
DB 482 CTTCCGATCTGTGTGACAGCGCTGGGCTCTGTGCTCCTCACCGTCTGTGCTGC 541
QY 159 ProSerAlaAlaSerIleAspAlaAlaLeuLysAlaGlyValAsnAsnGlyAsnPro 177
DB 542 CCACGTGCTGATCCCGAGTGTGTGACCAAGACCAATAAACCCTCGACCC 598

RESULT 12
BU447569 576 bp mRNA linear EST 29-NOV-2002
LOCUS 603764104F1 CSEQRBN13 Gallus gallus CDNA clone Chest687g2 5', mRNA
DEFINITION
SEQUENCE
VERSION BU447569
KEYWORDS BU447569.1 GI:25936880
SOURCE
ORGANISM Gallus gallus (chicken)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archaeopteryx; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 576)
REFERENCE
AUTHORS Boardman,P.B., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken CDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)

MEDLINE

2235534
PUBMED 12445392
CONTACT: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST
)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source

1..576

Location/Qualifiers

/organism="Gallus gallus"

/mol_type="mRNA"

/strain="Layer"

/db_xref="taxon:9031"

/clone="Chest687g2"

/sex="Female"

/dev_stage="adult"

/lab_host="DH10B"

/clone_lib="CSEQRBN13"

/note="Organ: ovary; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
, compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldi et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

BASE COUNT 150 a 169 c 145 g 112 t

ORIGIN

ALIGNMENT SCORES:
Pred. No.: 1.38e-05 Length: 576
Score: 146.00 Matches: 48
Percent Similarity: 46.94% Conservative: 21
Best Local Similarity: 32.65% Mismatches: 58
Query Match: 15.60% Indels: 20
DB: 13 Gaps: 8

US-09-589-870B-2 (1-183) x BU447569 (1-576)

QY 34 ValSerAlaAlaGluAlaGlyIleThrGlyThrTrpTyraGlnLeuGlySerThrPhe 53
DB 95 CTCTCTGCCAGAAAGTCTGCTGAGTGGAAATGAGCAACGATCTGGCTCCAAATG 154
QY 54 IleValThrAla---GlyAlaSerGlyAlaLeuThrGlyThrTyrgluserAlaValGly 72
DB 155 ACCATCGGGGCTGTGAACAGCAGAGTGAATTCACAGGACCTTACATCACACCGTA 211
QY 73 AsnAlaGluSerArgTyraValLeuThrGlyArgTyraSerAlaProAlaThrAspGly 92
DB 212 ACAGCCACATCAAAAGAGATCAAGATCACCATGATGAGACCAAAACCCATCAAC 271
QY 93 SerGlyThr-----AlaLeuGlyThrTrpThrValAlaTrpLysAsnAsnTyraGlnAla 110
DB 272 AATAGACCCACGCCACCTTGGCTTACCGGCAATTGGAG-----TTTTC 319
QY 111 HisSerAlaThrThrTrpSerGlyGlnTyraVal-----GlyAlaGluAlaArg 127
DB 320 GAGTCCACACCTGTCTTACACGGGCGGCTTCAATAGACAGAAATGGAGAGGTC--- 376
QY 128 IleAsnThrGlnTrpLeuLeuThrSerGlyThrGlu---AlaAsnAlaTrpLysSer 146
DB 377 CTGAAGACCATGTGCTGTGGGTCAAGTGAATGACATTGATGATGATGATGAAAGCT 436
QY 147 ThrLeuValGlyHisAspThrPheThrLysValLys-----Pro 159

D_b 437 ACCAGGCTCGGCATCAACACTTTTCACTCGCCCTGGCACACAGAAGCACTAGCATGGCCC |||||

Q_y 160 SerAlaIaseSerIleAspAla 166
 ||||| ::|||

D_b 497 CGCAAGCCGACACAATGCC 517

FEATURES	Location/Qualifiers
source	1. .622

BASE COUNT	108 a	210 c	180 g	124 t
ORIGIN				
Alignment Scores:				
Pred. No.:	1.73e-05	Length:	622	
Score:	145.50	Matches:	50	
Percent Similarity:	42.46%	Conservative:	26	
Best Local Similarity:	27.93%	Mismatches:	62	
Query Match:	15.54%	Indels:	41	
BB:	13	Gaps:	9	

Oy	34	ValSerAlaAlaGluValAlaGlyIleLeuThrGlyThrTrpTrpAsnGlnLeuGlySerThrPro	53
Db	71	GTGGCCCTCCAGAAAGTGGAGACTCCAGGCGCTTGAGAGAAACGACTGGCTCCAAACTG	130
Oy	54	IleValThrAla--GlyAlaAspGlyAlaLeuThrGlyThrTYrGluSerAlaValGly	72
Db	131	ACCACTCCAGCCCTGGATGGCTGGGACCCTTCCTCCGGCTCTTCCAGACAGCGGTG---	187
Oy	73	AsnIaGluSerArgTYrValLeuThrGlyArgTYrAspSerAla-----ProAla	89
Db	188	ACACCCACCAACAAGCAGATCTGTGTGACCCGCTGAAGGGCCACGACCCCGCGC	247
Oy	90	ThrAspGlySerGlyThrAlaLeuGlyTYrThrValAlaATrplysAsnAsnTYrArgAsn	109
Db	248	ACCAAGGGGGCAGCAGCCACCCTTTGGGCTTCACTGTGACAGTG-----CAGATT	295
Oy	110	AlaHisSerAlaThrTrpTrpSerGlyGln---TYrValGly-----GlyAlaGluAla	126
Db	296	GCAAGACTCCACCACTGCTCTTTGTGGAGACAGTCTTGTCGACCGCCGTGGAGAGAG--	352
Oy	127	ArgIleAsnThrGlnTrpLeuLeuThrSerGlyThr---ThrGluAlaAsnAlaTrplys	145
Db	353	ATGCTTGAGATGGCATGGTGGCTCGCGGGAAGAGTTCCCTCCGCCAAGACACCTGGAAAG	412
Oy	146	SerThrLeuValGlyHisAspThrPheThrTyrAllys	158
Db	413	GCACCCAGGGGTGGACCAATGTCTTCAACCGGTCAAGTGAATGATGGGAGACCTGGCTGG	472
Oy	158	-----	158
Db	473	CCTGCGCATCTGTCGTGAGGCGCTGGGACTCTGTGCTCATCACCGTGTCTGTGTC	532
Oy	159	ProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValaAsnAsnGlyAsnPro	177
Db	533	CCACTCTCTGTGATCCAGTGGCTGTGTGCACCAAGGACCAATATAAACCCCCCTGCACC	589

RESULT 14	
BU447766	
LOCUS	
DEFINITION	BU447766 575 bp mRNA linear EST 29-NOV-2002
ACCESSION	603221210P1 CSEQRBN13 Gallus gallus CDNA clone CHEST193K7 5' mRNA
VERSION	sequence.
KEYWORDS	BU447766 BU447766
ORGANISM	BU447766.1 GI:25937064
	EST.
	Gallus gallus (chicken)
	Gallus gallus
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
	Phasianinae; Gallus.
REFERENCE	1 (bases 1 to 575).
AUTHORS	Boardman,P.E., Sanz-Eguerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
TITLE	Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
JOURNAL	A Comprehensive Collection of Chicken CDNs
COMMENT	Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE	22335534
PUBMED	12445392
CONTACT	Contact: Simon Hubbard
	Department of Biomolecular Sciences
	University of Manchester Institute of Science and Technology (UMIST)
)
	PO Box 88, Manchester, M60 1QD, UK
	Tel: 01612008930
	Fax: 01612360409
	Email: Simon.Hubbard@umist.ac.uk.
	Location/Qualifiers
FEATURES	
source	1. 575

```

/1ab_host="DH10B"
/clone_1lb="CSEORBN3"
/Note="Organ: ovary; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI. This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methyated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."
BASE COUNT      148 a      170 c      146 g      111 t
ORIGIN
Alignment Scores:
Pred. No.:      1,766-05      Length:      575
Score:          145.00      Matches:      48
Percent Similarity: 46.26%      Conservative: 20
Best Local Similarity: 32.65%      Mismatches: 59
Query Match:     15.49%      Indels:      20
DB:              13      Gaps:      8
US-09-589-870B-2 (1-183) x BU447766 (1-575)
QY      34 ValSerAlaAlaGluAlaGlyIleThrGlyThrTyrAsnGlnLeuGlySerThrPhe 53
      100 CTTCTGCGCAGAAAGGCTCGTGAATGAGCAACGATCTGGCTCAACAG 159
QY      54 IleValThrAla---GlyAlaAspGlyAlaLeuThrGlyThrTyrGluSerAlaValGly 72
      160 ACCATCGGGCGCTGTGAACAGCAGGTAATTCACAGCGCCATCAACGCCGTA--- 216
DB      73 AsnAlaGluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAspGly 92
      217 ACAGCCACATCAATAGATCAAGATCAACGCCATGATGGACACAAACACCATCAAC 276
QY      93 SerGlyThr-----AlaLeuGlyTyrThrValAlaTyrPlyAsnAsnTyrArgAsnAla 110
      277 AAGAGAGCCAGCCAGCCACTTGGCTTCAACGCTCAATTGGAAG-----TTTTC 324
DB      111 HisSerAlaThrThrTyrSerGlyGlnTyrVal-----GlyGlyAlaGluAlaArg 127
      325 GAGTCCACCACTCTTCAACGCGCCAGTCTTCATAGACAGAAATGGGAAGAGGTC--- 381
QY      128 IleAsnThrGlnTyrPleuLeuThrSerGlyThrThrGlu---AlaAsnAlaTyrPlySer 146
      382 CTGAAGACCATGTGGCTGCTGCGCTCAAGTGAATGACTGCTGATGACTGGAAGACT 441
DB      147 ThrLeuValGlyHisAspThrPheThrValys-----Pro 159
      442 ACCAGAGTGGCATCAACATCTTCACTCGCTCGCACACAGAAAGATGAGATGGCC 501
QY      160 SerAlaAlaSerIleAspAla 166
      502 CGCAAGCCAGCAACATGCC 522
RESULT 15
LOCUS      BU408647      576 bp      mRNA      linear      EST 29-NOV-2002
DEFINITION 603158704F1 CSEORBL03 Gallus gallus cDNA clone Chest177g1 5', mRNA
sequence.
ACCESSION  BU408647
VERSION    BU408647.1 GI:25901318
KEYWORDS   EST
SOURCE     Gallus gallus (chicken)
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

```

```

Phasianidae; Gallus.
1 (bases 1 to 576)
AUTHORS    Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Boech, E.,
Fong, W. T., Tickle, C., Brown, W. R. A., Wilson, S. A. and Hubbard, S. J.
TITLE      A Comprehensive Collection of Chicken cDNAs
JOURNAL    Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE    22335534
PUBMED     12445392
COMMENT    Contact: Simon Hubbard
            Department of Biomedical Sciences
            University of Manchester Institute of Science and Technology (UMIST)
            PO Box 88, Manchester, M60 1QD, UK
            Tel: 01612008930
            Fax: 01612260409
            Email: Simon.Hubbard@umist.ac.uk.
FEATURES   source
            1..576
            /organism="Gallus gallus"
            /mol_type="mRNA"
            /strain="Layer and broiler"
            /db_xref="taxon:9031"
            /clone="Chest177g1"
            /sex="Male and female"
            /tissue_type="Chondrocytes isolated from growth plate cartilage"
            /dev_stage="adult"
            /lab_host="DH10B"
            /clone_1lb="CSEORBL03"
            /note="Vector: pBluescript II KS(+); Site_1: EcoRI;
            Site_2: NotI. Modification of pBluescript II KS(+)
            (Stratagene) vector to accommodate cDNA produced with the
            trimmed protocol (construction of uni-directionally
            cloned cDNA libraries from messenger RNA for improved 3'
            end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387
            ,624). Cut pBluescript II KS(+) with NotI and EcoRI.
            ligate in double stranded adaptor containing BamI and
            BamH sites [5'ggcgcgtgcagcccgagccggaataaag]
            [5'aattcttttcggatccgggctcgacgc]"
BASE COUNT      148 a      170 c      147 g      111 t
ORIGIN
Alignment Scores:
Pred. No.:      1,766-05      Length:      576
Score:          145.00      Matches:      48
Percent Similarity: 46.26%      Conservative: 20
Best Local Similarity: 32.65%      Mismatches: 59
Query Match:     15.49%      Indels:      20
DB:              13      Gaps:      8
US-09-589-870B-2 (1-183) x BU408647 (1-576)
QY      34 ValSerAlaAlaGluAlaGlyIleThrGlyThrTyrAsnGlnLeuGlySerThrPhe 53
      95 CTTCTGCGCAGAAAGGCTCGTGAATGAGCAACGATCTGGCTCAACAG 154
DB      54 IleValThrAla---GlyAlaAspGlyAlaLeuThrGlyThrTyrGluSerAlaValGly 72
      155 ACCATCGGGCGTGTGAACAGCAGGTAATTCACAGCGCCATCAACGCCGTA--- 211
QY      73 AsnAlaGluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAspGly 92
      212 ACAGCCACATCAATAGATCAAGATCAACGCCATGATGGACACAAACACCATCAAC 271
DB      93 SerGlyThr-----AlaLeuGlyTyrThrValAlaTyrPlyAsnAsnTyrArgAsnAla 110
      272 AAGAGAGCCAGCCAGCCACTTGGCTTCAACGCTCAATTGGAAG-----TTTTC 319
QY      111 HisSerAlaThrThrTyrSerGlyGlnTyrVal-----GlyGlyAlaGluAlaArg 127
      320 GAGTCCACCACTGTTCACGCGCCAGTCTTCATAGACAGAAAGATGAGATGGCC 376
DB      128 IleAsnThrGlnTyrPleuLeuThrSerGlyThrThrGlu---AlaAsnAlaTyrPlySer 146

```

```

Db      377 CTGAAGACCATGTGCTGCTGCGCAAGTGTATGACATGCTGATGACTGGAAGCT 436
Oy      147 ThrlauValGlyHisAspThrPheThrLysValLys-----Pro 159
Db      437 ACCAGGCTCGGCATCAACATCTTCACTGCGCACACAGAGAGAGTGAAGATGGCC 496
Oy      160 SerAlaIaSerIleAspAla 166
Db      497 CGCAAGCCGACACATGCC 517

```

Search completed: October 28, 2003, 02:05:33
 Job time : 1939 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: October 28, 2003, 00:19:43 ; Search time 253 Seconds
(without alignments)
1952.558 Million cell updates/sec

Title: US-09-589-870b-2
Perfect score: 936
Sequence: 1 MKRIVVAIAIVSLTVSITA.....IDAAKAGVNGNPNLDVQO 183

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_plus.model -DEV=xlh
-Q=/cgn2_1/USPTO.epool/US05989870/runat_27102003_104445_16575/app_query.fasta_1.327
-DB=N_Geneseq_19Jun03 -QPMT=fastag -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US05989870.@CGN_1_1.312.@runat_27102003_104445_16575 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DBPLOCK=100 -LONLOCK
-DB TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
- 25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	936	100.0	552	20	AAK80198	Streptococcus stre
2	936	100.0	563	21	AAZ91073	Streptomyces avidi
3	936	100.0	638	18	AAT73193	Wild-type streptav
4	936	100.0	638	19	AAV34714	S. avidinii strept
5	936	100.0	638	21	AAZ49867	Streptavidin gene
6	936	100.0	638	22	AAC86561	DNA encoding a str
7	936	100.0	1131	7	AAK60626	Sequence of a port
8	931	99.5	564	21	AAZ91074	Streptomyces avidi
9	927	99.0	552	14	AAQ53412	Streptavidin gene.
10	919	98.2	638	19	AAV34715	S. avidinii strept
11	917	98.0	638	19	AAV34716	S. avidinii strept
12	855.5	91.4	1239	22	AAC86563	DNA encoding a fus
13	854.5	91.3	1612	22	AAC86562	DNA encoding a fus
14	847	90.5	1280	22	AAC86564	Anti-CD20 single c
15	846	90.4	1356	15	AAQ70659	Scfv PRAS108 and p
16	838	89.5	495	21	AAZ91076	Streptomyces avidi
17	835	89.2	492	21	AAZ91075	Streptomyces avidi
18	835	89.2	498	9	AAK80186	Encodes biosynthes
19	834	89.1	498	18	AAT75491	DNA for streptavid
20	834	89.1	507	10	AAK90755	Sequence of strept
21	834	89.1	1266	24	ABA92275	Streptavidin-carbo
22	834	89.1	7306	25	ABZ77370	Nucleotide sequenc
23	834	89.1	8155	25	ABZ77369	Nucleotide sequenc
24	772	82.5	1881	25	AAQ50166	Biotin-combining s
25	745	79.6	1296	15	AAQ70660	Scfv PRAS109 and p
26	687.5	73.5	1173	18	AAT86312	Single-chain anti-
27	666	73.3	864	20	AAK89302	S. avidinii protein
28	682	72.9	1175	18	AAT86311	Single chain anti-
29	679	72.5	384	24	ABZ61154	S. avidinii strept
30	679	72.5	401	21	AAZ49868	Streptavidin cDNA f
31	679	72.5	2019	17	AAT33850	DNA encoding strep
32	675	72.1	384	18	AAT73195	Recombinant Core-s
33	675	72.1	387	14	AAQ40718	Core streptavidin.
34	673.5	72.0	1257	15	AAQ70661	Scfv PRAS110 and p
35	672	71.8	626	21	AAZ49866	Potato proteinase
36	660	70.5	2055	17	AAT33851	DNA encoding mutan
37	658.5	70.4	1247	24	ABK12869	Chimeric SA-FastL D
38	656	70.1	1266	24	ABK12871	Chimeric TGF-B-SA
39	656	70.1	1329	24	ABK12872	Chimeric IL-2-SA D
40	656	70.1	1395	24	ABK12873	Chimeric IL-10-SA
41	656	70.1	1542	24	ABK12870	Chimeric B7.1-SA D
42	656	70.1	1566	24	ABK12874	Chimeric SA-CD40L
43	650	69.4	369	21	AAZ91077	Streptomyces avidi
44	650	69.4	376	21	AAZ91078	Streptomyces avidi
45	644	69.8	466	25	ABZ70370	Streptavidin-based

ALIGNMENTS

RESULT 1
AAK80198
ID AAK80198 standard, cDNA, 552 BP.
AC AAK80198;
XX
XX 20-AUG-1999 (first entry)
XX
XX Streptococcus streptavidin encoding cDNA.
DE
XX Avidin; streptavidin; batroxobin; fibrinogen converting enzyme;
KW hybrid; fusion protein; sealant; surgery; reduce bleeding; fibrin; ss.
XX
XX Streptococcus sp.
OS
XX
XX
XX WO9929838-A1.
XX


```

US-09-589-870B-2 (1-183) x AA291073 (1-563)
QY 1 MetArgLysIleValValAlaAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
DB 10 ATGCGCAAGATCGTCTGTCGACGATCGCGCTTCCCTGACCAAGGCTCTGATTACGGCC 69
QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaGly 40
DB 70 AGCGCTTCGGCAACCCCTCCAAAGACTCGAAGGCCAGGTCTTCGGCCGGCGGCGCGC 129
QY 41 IleThrGlyThrTyrPheGlnLeuGlySerThrPheIleValThrAlaGluAlaAsp 60
DB 130 ATCAGCGGACACTGTGATCAACAGCTCGGCTCCAGCTTCACTGACCGCGGCGCGCAG 189
QY 61 GlyAlaLeuThrGlyThrTyrGluSerAlaValGlyAlaAsnAlaGluSerArgTyrValLeu 80
DB 190 GGGGCGCTGACCGGAACCTACGAGTCGGCGCGGCAACCGCCGAGAGCCGCTACGCTCG 249
QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
DB 250 ACCGGTCGTTACAGACGCGCCCGGACCGGACGCGGACCGCGCTCGGTTGGAGC 309
QY 101 ValAlaTyrLysAsnAsnTyrArgAsnAlaIleSerAlaThrThrTyrPheSerGlyGlnTyr 120
DB 310 GTGGCGTGGAGAATACTACCGCAACGCCACTCCGCGGACGATGAGCGGCGGCGGAGTAC 369
QY 121 ValGlyGlyAlaGluAlaArgIleAsnThrGlnTyrLeuLeuThrSerGlyThrThrGlu 140
DB 370 GTGGCGGCGCGCGGAGGATCAACACCCAGTGTGCTGACCTCCGCGCACCGACCGAG 429
QY 141 AlaAsnAlaTyrLysSerThrLeuValGlyAlaAspThrPheThrLysValLysProSer 160
DB 430 GCCAAGCGCTCGAAGTCTCAGCTGCTGCGGACGACACCTTCCAAAGTGAAGCCGCTCC 489
QY 161 AlaAlaSerIleAspAlaAlaLysLysAlaGlyValAlaAsnGlyLysProLeuAspAla 180
DB 490 GCCGCGTCCATCGACGCGGCGGAAGAGCGCGGTCAACAACGCGCTCGACCGCC 549
QY 181 ValGlnGln 183
DB 550 GTTCAGCAG 558

RESULT 3
AAT73193
ID AAT73193 standard; DNA; 638 BP.
AC AAT73193;
XX
DT 27-APR-1998 (first entry)
DE Wild-type streptavidin DNA.
XX
KM Streptavidin; biotin; anti-interference reagent; detection; muten;
KM avidin; non-specific binding; ss.
XX
OS unidentified.
XX
FH Key Location/Qualifiers
FT CDS 50..601
FT FT /*tag= a
FT FT /product= streptavidin
FT FT sig_peptide 50..121
FT FT /*tag= b
FT FT mat_peptide 122..598
FT FT /*tag= c
FT FT /product= streptavidin
XX
PN DE19637718-A1.
XX
PD 02-OCT-1997.
XX
PF 16-SEP-1996; 96DE-1037718.

```

```

XX
PR 01-APR-1996; 96DE-1013053.
XX
PA (BOER ) BOEHRINGER MANNHEIM GMBH.
XX
PI Brandstetter H, Deger A, Engh R, Kopetzki E, Mueller R;
PI Schmitt U;
XX
DR WPI: 1997-482043/45.
DR P-PSDB; AAM29306.
XX
PT Streptavidin and avidin muteins with reduced binding affinity for
PT biotin - useful for reducing interference from nonspecific binding
PT in assays
XX
PS Disclosure; Page 16-17; 26pp; German.
XX
CC This DNA sequence encodes a streptavidin which is used in a novel method
CC of reducing interference from non-specific binding in assays. Muteins
CC constructed from a core streptavidin or avidin sequence are selected that
CC differ from the native polypeptide by at least one amino acid and have a
CC binding affinity for biotin of less than 1010 1/mole. The biotin-bindable
CC polypeptide may be present as a polymeric conjugate, e.g. with another
CC polypeptide or protein, especially bovine serum albumin. These muteins
CC are used as anti-interference reagents for reducing and/or avoiding
CC nonspecific interactions in a process for detecting an analyte. In
CC particular, they are used in assays where the streptavidin/avidin-biotin
CC specific binding pair is involved for qualitative and/or quantitative
CC determination of an analyte in a test sample, e.g. a heterogeneous
CC immunoassay or a hybridisation assay. Despite having a lower binding
CC affinity for biotin, the muteins have high immunological cross-reactivity
CC with native streptavidin and avidin.
XX
SQ Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;

Alignment Scores:
Pred. No.: 5, 93e-74 Length: 638
Score: 936.00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-589-870B-2 (1-183) x AAT73193 (1-638)
QY 1 MetArgLysIleValValAlaAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
DB 50 ATGCGCAAGATCGTCTGTCGACGATCGCGCTTCCCTGACCAAGGCTCTGATTACGGCC 109
QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaGly 40
DB 110 AGCGCTTCGGCAACCCCTCCAAAGACTCGAAGGCCAGGTCTCGGCCCGCGGCGCGC 169
QY 41 IleThrGlyThrTyrPheGlnLeuGlySerThrPheIleValThrAlaGluAlaAsp 60
DB 170 ATCAGCGGACACTGTGATCAACAGCTCGGCTCCAGCTTCACTGACCGCGGCGCGCAG 229
QY 61 GlyAlaLeuThrGlyThrTyrGluSerAlaValGlyAlaAsnAlaGluSerArgTyrValLeu 80
DB 230 GGGGCGCTGACCGGAACCTACGAGTCGGCGCGGCAACCGCGAGAGCGCTACGCTCG 289
QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
DB 290 ACCGGTCGTTACAGACGCGCCCGGACCGGACGCGGACCGCGCTCGGTTGGAGC 349
QY 101 ValAlaTyrLysAsnAsnTyrArgAsnAlaIleSerAlaThrThrTyrPheSerGlyGlnTyr 120
DB 350 GTGGCGTGGAGAATACTACCGCAACGCCACTCCGCGACGATGAGCGGCGGCGGAGTAC 409
QY 121 ValGlyGlyAlaGluAlaArgIleAsnThrGlnTyrLeuLeuThrSerGlyThrThrGlu 140
DB 410 GTGGCGGCGCGGAGGAGATCAACACCCAGTGTGCTGACCTCCGCGCACCGAGCAG 469

```

QY 141 AlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrIysValLysProSer 160
 Db 470 GCCAACCCCTGGAAGTCCAGCTGTCGGCCACGACCTTCAACGAAGTGAAGCCGTC 529
 QY 161 AlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsnGlyAsnProLeuAspAla 180
 Db 530 GCCGCTCCATCGACGGCGGGAAGAGCCGGCGTCAACACGCAACCCGCTCGACGCC 589
 QY 181 ValGlnGln 183
 Db 590 GTTCAGCAG 598
 RESULT 4
 AAV34714
 ID AAV34714 standard; DNA; 638 BP.
 XX
 AC AAV34714;
 XX
 DT 27-AUG-1998 (first entry)
 XX
 DE S. avidinii streptavidin cDNA.
 XX
 KM Streptavidin; ligand; binding affinity; mutant; isolation;
 XX purification; recover; immobilise; ss.
 OS Streptomyces avidinii.
 XX
 FH Key Location/Qualifiers
 FT CDS 50..601
 FT sig_peptide /*tag= a
 FT 50..121
 FT mat_peptide /*tag= b
 FT 122..598
 FT /*tag= c
 FT /product= Streptavidin
 XX
 PN EP835934-A2.
 PD 15-APR-1998.
 XX
 PF 09-OCT-1997; 97EP-0117504.
 XX
 PR 10-OCT-1996; 96DE-1041876.
 XX
 PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
 XX
 PI Skerra A, Voss S;
 XX
 DR WPI: 1998-218668/20.
 DR P-PSDB; AAMS9216.
 XX
 PT Streptavidin mutants with higher binding affinity for peptide
 PT ligands - have mutation in amino acid region 44-53, used to isolate,
 PT purify or determine fusion proteins including these ligands
 XX
 PS Disclosure; Page -; 21pp; German.
 XX
 CC This sequence encodes a wild-type streptavidin protein isolated from
 CC Streptomyces avidinii. This sequence is used to produce mutants which
 CC are used in a method to assay the binding affinity of streptavidin
 CC mutants. These mutants have a mutation within the amino acid (aa) region
 CC 44-53 of the wild-type protein show a higher binding affinity than the
 CC wild-type for peptide ligands that include the sequence of formula
 CC Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z are both Gly,
 CC or Y = Glu and Z = Arg or Lys. Recombinant streptavidin mutants can
 CC be used to isolate, purify and determine proteins or to
 CC determine/recover substances that contain streptavidin-binding groups.
 CC Such compounds may also be used to immobilise fusions on microtitre
 CC plates, microbeads or sensor chips.
 CC NOTE: This sequence does not appear in the specification but is used to
 CC make the mutant streptavidin sequence represented in AAV34715 and
 CC AAV34716.

SQ Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;
 Alignment Scores:
 Pred. No.: 5 93e-74 Length: 638
 Score: 936.00 Matches: 183
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 19 Gaps: 0
 US-09-589-870B-2 (1-183) x AAV34714 (1-638)
 QY 1 MetArgLysIleValAlaAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
 Db 50 ATCCGCAAGATCGTGTTCGACGACCATCGCCGTTTCCCTGACCAACGCGTCCGATTCGGCC 109
 QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGlnAgl 40
 Db 110 AGCGCTTCGGAGACCCCTCCCAAGAGCTCGAAGGCCAGGTCTCCGCGCCGAGGCCGCGC 169
 QY 41 IleThrGlyThrTrpTrpAsnGlnLeuGlySerThrPheIleValThrAlaGlyAlaAsp 60
 Db 170 ATCACCGGACACTGCTGACACCAAGCTCGCTGACCTTCATCTGATGACCGGCGCCGAC 229
 QY 61 GlyAlaLeuThrGlyThrTyrgLysSerAlaValGlyAsnAlaGluSerArgTyValLeu 80
 Db 230 GCGCGCTCGACCGGAACCTAGACAGTCCGCGCCGCAACGCGAGAGCGGCTAGCTCTG 289
 QY 81 ThrGlyArgTyAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
 Db 290 ACCGGTCTGACGAGAGCGCCCGCCGACCGACCGGACCGGACCGCCCTCGTTGAGC 349
 QY 101 ValAlaTrpLysAsnAsnTyArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyr 120
 Db 350 GTGGCTCGAAGAAATCACTACCGCAACCCCACTCCGACCACTGAGAGCGCCAGTAC 409
 QY 121 ValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGlu 140
 Db 410 GTCCGCGCGCGCGAGGATCAACACCAAGTGGTGTGACTGACCTCCGACCAACCGAG 469
 QY 141 AlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrIysValLysProSer 160
 Db 470 GCCAACCCCTGGAAGTCCAGCTGTCGGCCACGACCTTCAACGAAGTGAAGCCGTC 529
 QY 161 AlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsnGlyAsnProLeuAspAla 180
 Db 530 GCCGCTCCATCGACGGCGGGAAGAGCCGGCGTCAACACGCAACCCGCTCGACGCC 589
 QY 181 ValGlnGln 183
 Db 590 GTTCAGCAG 598
 RESULT 5
 AAZ49867
 ID AAZ49867 standard; cDNA; 638 BP.
 XX
 AC AAZ49867;
 XX
 DT 25-APR-2000 (first entry)
 XX
 DE Streptavidin gene for recombinant pART27 vector.
 XX
 KM Potato proteinase inhibitor-II; PPI-II; streptavidin; worm;
 KM insect; plant-noxious protein; pest resistance; moth; insect; weevil;
 KM grub; beetle; fly; thrip; locust; cricket; borer; mite; looper;
 XX insecticidal; ss.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 50..601
 FT /*tag= a
 FT /product= "Streptavidin protein"

```

FT      sig_peptide      50..121
FT      /*tag= b
FT      mat_peptide      122..598
FT      /*tag= c
PN      WO200004049-A1.
XX      27-JAN-2000.
XX      15-JUL-1999; 99MO-NZ00110.
XX      15-JUL-1998; 98NZ-0331002.
XX      (HORT-) HORTICULTURE & FOOD RES INST NEW ZEALAND.
PI      Christeller JT, Sutherland PW, Murray C, Markwick NP, Phillip BA,
PI      Malone LA, Burgess EPJ;
DR      MPI; 2000-171244/15.
DR      P-PSDB; AAY44701.
XX      PT      New chimeric polypeptide and composition comprising the polypeptide
XX      PT      useful for conferring pest resistance on plants
XX      PS      Disclosure; Fig 12; 11pp; English.
XX      CC      The present sequence encodes streptavidin, a plant-noxious protein. This
XX      CC      is used in the preparation of a recombinant vector, pAKR27 designed to
XX      CC      express a chimeric polypeptide comprising streptavidin mature peptide
XX      CC      fused to the potato proteinase inhibitor-II (PPI-II) signal peptide. The
XX      CC      binary vector is targeted to the vacuole by PPI-II signal sequence.
XX      CC      Transformation of plant genome with the vector can produce pest
XX      CC      resistance in plants, plant derived products and stored harvest
XX      CC      material. Pests that can be controlled include, cotton bollworm,
XX      CC      tropical army-worm, European corn-borer or red mite, tobacco horn worm,
XX      CC      Colorado potato beetle, European spruce sawfly, diamondback moth, potato
XX      CC      tuber moth, codling moth, Indian meal moth, gypsy moth, Argentine stem
XX      CC      weevil, clover root weevil, grass-grubs, corn rootworm, rice and wheat
XX      CC      weevils, mealworms, flour beetles, black field cricket, locusts,
XX      CC      sawflies, Western flower thrips, Hessian flies or two-spotted mite.
XX      SQ      Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;

Alignment Scores:
Pred. No.: 5,93e-74 Length: 638
Score: 936.00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-589-870B-2 (1-183) x AA249867 (1-638)
QY      1 MetArgLySerIleValAlaAlaAlaIleAlaValSerIleuThrValSerIleThrAla 20
DB      50 ATGCGCAAGATCGTCTTGACGACCATCGCGTTTCCCTGACCGGTCTCGATTACGGCC 109
QY      21 SerAlaSerAlaAspProSerIleuAspSerIleuAlaGlnValSerAlaAlaGluAlaGly 40
DB      110 AGGCGTTGCGGACAGCCCTCCAGACTCGAAGGCCAGGTCTCGCGCCGCGAGCGCGCG 169
QY      41 ILeThrGlyThrTrpIleuValGlnIleuValSerIleuThrPheIleValThrAlaGlyAlaAsp 60
DB      170 ATTCACGGGACCGCTGACCAACGCTGCGCTCATCTTATCTGACCGCGCGCGCGAC 229
QY      61 GlyAlaLeuThrGlyThrIleuValGlnIleuValSerIleuThrValSerIleuThrVal 80
DB      230 GGGGCGCTGACCGGAACTACGAGTGGCGCTGCGGACCGCGGACCGCTGACGCTG 289
QY      81 ThrGlyArgTrpAspSerAlaProAlaIleuValSerIleuThrValSerIleuThrVal 100
DB      290 ACCGCGTGTACGACAGCGCGCGCGGACCGAGCGGACCGCGCGCTGCGTGGACG 349

```

```

QY      101 ValAlaTrpIleuAspAsnTrpArgAlaAlaIleSerAlaThrThrTrpSerGlyGlnTrp 120
DB      350 GTGGCTGGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 409
QY      121 ValGlyGlyAlaGlyAlaAlaGlyIleAsnThrGlnTrpIleuThrSerGlyThrGln 140
DB      410 GTGGGCGGCGCGGAGGATCAACACCGAGTGTGTGATCACTCCGCGACCAACCGAG 469
QY      141 AlaAsnAlaTrpIleuValGlyIleAspTrpPheThrIleValIleProSer 160
DB      470 GCCAAGCGCTGAGTCCAGCTGCTGCGCGCACACCTTACCAAGGTGAAGCGTCC 529
QY      161 AlaAlaSerIleAspAlaAlaIleValIleValIleValIleValIleValIleValIle 180
DB      530 GCGGCTTCATGACGCGCGGAGGAGGCGGCTGACACGACGCAACCGCTCGAGCGC 589
QY      181 ValGlnGln 183
DB      590 GTTCAGCAG 598

RESULT 6
AAC86561
ID      AAC86561 standard; DNA; 638 BP.
XX      AC      AAC86561;
XX      DT      02-APR-2001 (first entry)
XX      DE      DNA encoding a streptavidin polypeptide.
XX      KW      Streptavidin; tumour cell; cancer; adenocarcinoma;
XX      OS      hematological malignancy; ss.
XX      ST      Streptomyces avidinii.
XX      FH      Key Location/Qualifiers
XX      FT      CDS 50..601
XX      FT      /*tag= a
XX      FT      /product= "streptavidin"
XX      FT      sig_peptide 50..121
XX      FT      /*tag= b
XX      PN      WO200075333-A1.
XX      PD      14-DEC-2000.
XX      PF      05-JUN-2000; 2000WO-US15595.
XX      PR      07-JUN-1999; 99US-0137900.
XX      PR      03-DEC-1999; 99US-0168976.
XX      PA      (NEOR-) NEORX CORP.
XX      PI      Goshorn SC, Graves SS, Schultz JE, Lin Y, Sanderson JA, Reno JW;
XX      PI      MPI; 2001-091213/10.
XX      DR      P-PSDB; AAB30692.
XX      PT      New vector constructs for expressing genomic streptavidin fusion
XX      PT      proteins which are useful for targeting tumour cells associated with
XX      PT      cancer, e.g. adenocarcinomas
XX      PS      Claim 14; Fig 4; 100pp; English.
XX      CC      The present sequence encodes a streptavidin polypeptide. The sequence
XX      CC      is used to construct vectors of the invention. The specification
XX      CC      describes vector constructs for expressing streptavidin fusion proteins.
XX      CC      The vector comprises a first nucleic acid encoding genomic streptavidin
XX      CC      or its functional variant operatively linked to a promoter, and a
XX      CC      cloning site for insertion of a second nucleic acid sequence encoding a
XX      CC      polypeptide to be fused with streptavidin, interposed between the
XX      CC      promoter and the first nucleic acid sequence. Alternatively, the vector
XX      CC      construct comprises a first nucleic acid, operatively linked to a

```

CC promote encoding a polypeptide to be fused with streptavidin, and a
CC cloning site for insertion of a second nucleic acid encoding at least
CC 129 amino acids of streptavidin or its functional variant. The fusion
CC proteins are useful for targeting tumour cells, particularly tumour cells
CC associated with cancer, e.g. adenocarcinomas or hematological
CC malignancies. The vector construct is useful for expressing of
CC streptavidin fusion proteins. In particular, these are useful as tools
CC for medical diagnostics and therapeutic purposes, e.g. for detecting the
CC presence or absence of, or treating, a target site within a mammalian
CC host.

Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;

Alignment Scores:

Pred. No.:	5.93e-74	Length:	638
Score:	936.00	Matches:	138
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0

US-09-589-870B-2 (1-183) x AAC86561 (1-638)

Oy	1	MetArgLysIleValAlaIleAlaIleAlaIleValSerLeuThrThrValSerIleThrAla	20
Db	50	ATGGCAAGATCGTCGTTGCAAGCATCGCGTTTCCCTTACACACGGCTTCGATTACGGCC	105
Oy	21	SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGlnAlaGly	40
Db	110	AGCCCTTCGGAGACCCCTCCAAAGACTCGAAGGCCCAAGGTCTCGGCGCGAGCGCGGC	165
Oy	41	IleThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAlaAsp	60
Db	170	ATCACCGGCACCTGGTACACCGCTGGCTGCACTTCATGTACCGCGGCGCCGAC	225
Oy	61	GlyAlaIleuThrGlyThrTyrGlnSerAlaValAlaGlyAsnAlaGlnSerArgTyrValLeu	80
Db	230	GGCGCCCTGACCGGAACCTACGAGTGGCGCGTGGCAAGCCGGAAGCGCGCTACGCTCTG	285
Oy	81	ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr	100
Db	290	ACCGGTGTTACACACAGCGCGCCCGCCACCGACGGAGCGGACCGCCCTGGTTGACG	345
Oy	101	ValAlaIleTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyr	120
Db	350	GTGGCCTGGAAAGATTACTACCCCAACGCCCACTCCGACACACGTGAGAGCGGCACATAC	405
Oy	121	ValGlyGlyAlaGlnAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGln	140
Db	410	GTCGGCGCGCCCGAGGGAGATCAACACCCAGTGGCTGTGACCTCGGACACCAACCGAG	465
Oy	141	AlaAsnAlaIleTrpLysSerThrLeuValGlyHisAspThrPheThrLysValLysProSer	160
Db	470	GCCACGCCCTCGAAAGTCCACGCTGGTCGCCACGACACCTTCACCAAGGTGAACCGCTCC	525
Oy	161	AlaAlaSerIleAspAlaAlaLysLysAlaGlyValAlaAsnGlyAsnProLeuAspAla	180
Db	530	GCGCGCTCCATCGACCGCGGAAGAGCGCGCGTCAACAAAGCGAACCCGCTTCAGCGCC	585
Oy	181	ValGlnGln 183	
Db	590	GTTCACGACG 598	
<p>RESULT 7 AAN60626 ID AAN60626 standard; DNM; 1131 BP. XX XX AAN60626; XX XX 25-MAR-2003 (updated) DT 13-AUG-1991 (first entry) XX XX Sequence of a portion of SA307 which encodes a streptavidin-like</p>			

DE polypeptide.
xy

Antibiotic; biotin binding affinity; fusion protein; ds.

OS Streptomyces
XY

FH	Key	Location/Qualifiers
ET	CNS	481 1000

```

/*tag= a
/not=
//AAB00000//

```

CDS	482..1030
ET	/++cc-
ET	b

```
ET /note= "AAP60624"
ET 480 1031
CDS
```

```
/*tag= c  
/proc/  
ET  
ET
```

FT /note= "AAP60625"
XX

PN MO8602077-A.
XX

FD 10-APR-1986.
XX

01-OCT-1985; 85WFO-0001901.
XX

02-OCT-1984; 84US-0656873.

PA (MEAD/) MEADE H M.
PA (BIOJ) BIOGEN NV.

Meade HM, Garwin JL, Biogen NV:

WPI: 1986-106643/16.

XX
DN F-FODB; HAF800623, HAF800624, HAF80
XX

streptavidin-like polypeptide, all

e.g. tissue plasminogen activator
XX

Discipline, 199, 2; 399p; English
XX

steptavidin-like polypeptide (see

and a second sequence coding for a

CC (Updated on 25-MAR-2003 to correct

SQ Sequence 1131 BP; 199 A; 409 C; 35

Alignment Scores:

Score: 936.00

Best Local Similarity: 100.00%

DB: 7

US-09-589-870B-2 (1-183) X AAN60626 (1-

1 MetArglysIleValValAlaAlaI

Db 480 ATGCGCAGATCGTCGTTGCAGCCF

21 SerAlaSerAlaAspProSerLysP
QY

Db 540 AGCGCTTGCGGACAGACCCCTCCAAAGCACTCGAAGCGCCCAAGTCTCGCGCCGCGAGCGCGC 59

QY 41 ILETRGJYThTtRPTYAsnGlnenglyserThrPheIleValThRAgIYlaaSP 60

QY 600 ATCAGCGGCACCTGGTACCAACGAGTCCGGCTCGACCTTCATGTATCCGCGGCGCCGCAC 65

Db 61 GJYAlaLeuThnGJYThTtYrGluSerIaValGJYAsnIaGluSerArgTYrValIeu 80

Db 660 GGGCCCTGACGGAACTACGATCGCGCCGCGCAACCGCAGACCGCTACGTCTTG 719
Qy 81 ThrGlyArgThrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTTPThr 100
Db 720 ACCGGTGTTCACACAGCCGCCGCCGACGCGGCGCGCCGCTCGGTGGAGC 779
Qy 101 ValAlaTrpIleAsnGlnTyrAsnAlaIleSerAlaThrThrTrpSerGlyGlnTyr 120
Db 780 GTGGCTGGAAAGATACTACCCGCAAGCCCATCTCCGCGACCACTGGAGCGCCAGTAC 839
Qy 121 ValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrGln 140
Db 840 GTGGCGGCGCCGAGGAGATCAACACCCAGTGGCTGTGACTCCGCGACCCAGC 899
Qy 141 AlaAsnAlaTrpIleSerThrLeuValGlyHisAspThrThrIleValIleProSer 160
Db 900 GCCAAGCGCTGGAAGTCAAGCTGGTGGCCAGCACCTTCAACAGGAGGCCGCTCC 959
Qy 161 AlaAlaSerIleAspAlaAlaIleValAlaGlyValAsnAsnGlyAsnProLeuAspAla 180
Db 960 GCGGCTCCATCCACGCGCGGAAGAGCCGCGCTCAACACCGCAACCCGCTCGAGCC 1019
Qy 181 ValGlnGln 183
Db 1020 GTTCAGCAG 1028
RESULT 8
AA291074
ID AA291074 standard; DNA: 564 BP.
XX
AC AA291074;
XX
DT 06-JUN-2000 (first entry)
XX
DE Streptomyces avidinii mat gene.
XX
KM plant somatic tissue degeneration; plant essential factor; depletion;
KM viability; mat gene; plant development; plant morphology; flower;
KM fruit plant; de.
XX
OS Streptomyces avidinii.
XX
PN MO200007427-A2.
XX
PD 17-FEB-2000.
XX
PF 30-JUL-1999; 99MO-IL00420.
XX
PR 03-AUG-1998; 98IL-0125632.
XX
PA (AGRI-) AGRIC RES ORG.
XX
PI Kapulnik Y, Ginzberg I;
XX
DR MPI; 2000-195402/17.
XX
DR P-RSDB; AA050513.
XX
PT Degeneration of somatic plant tissue by expression of a heterologous
PT protein, useful for controlling plant development and morphology, such
PT as decreasing the number of flowers present to increase the number of
PT fruit -
XX
PS Examples: Page 85; 91pp; English.
XX
CC The invention relates to a method of effecting degeneration of a somatic
CC plant tissue by expressing a heterologous protein capable of binding a
CC plant essential factor (PEF), in somatic plant tissue cells, where
CC heterologous protein expression causes depletion of the PEF so the plant
CC viability is maintained, while simultaneous degeneration of the somatic
CC plant tissue is effected. This sequence represents the Streptomyces
CC avidinii mat gene as an example of a heterologous gene introduced into
CC the plants. The methods can provide for the selective and optionally
CC reversible cell degeneration in somatic plant tissue. They can be used

CC for artificially controlling plant development and morphology. They can
CC be used e.g. to decrease the number of flowers in fruit producing plants
CC so as to increase the number of fruits which reach maturity.
XX
SQ Sequence 564 BP; 109 A; 213 C; 169 G; 73 T; 0 other;
Alignment Scores:
Pred. No.: 1,42e-73 Length: 564
Score: 931.00 Matches: 182
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.47% Indels: 0
DB: 21 Gaps: 0
US-09-589-870B-2 (1-183) x AA291074 (1-564)
Qy 2 ArgIleValIleValAlaIleAlaIleValIleSerLeuThrThrValSerIleThrAlaSer 21
Db 14 CCGAAGATGCTGCTGGACCATCCGCTTCCCTGACCAACGGTCCATTACGGCCAGC 73
Qy 22 AlaSerAlaAspProSerIleAspSerIleValGlnValSerAlaIleGluAlaGlyIle 41
Db 74 GCTTCGCGACACCCCTCCAAAGACTCGAAGGCCCAAGTCTCGGCGCCGAGGCGCGCATC 133
Qy 42 ThrGlyThrTrpTrpAsnGlnLeuGlySerThrIleValIleValIleAspGly 61
Db 134 ACCGGACCTGGTACACCAAGCTCGCTGACCTTCACTGACCGCGGCGCGCGAGCGC 193
Qy 62 AlaLeuThrGlyThrTyrGluSerAlaValGlyAsnAlaGluSerArgTyrValLeuThr 81
Db 194 GCCCTGACCGGAACCTAGACAGTCCGGCGTGGCAACCGGAGAGCGGCTACGCTGACC 253
Qy 82 GlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTTPThrVal 101
Db 254 GGTCTTACGACAGCGCCCGCCGCAACCGCAGCGGCGGACCGCCCTCGTTGAGCGTG 313
Qy 102 AlaTrpIleAsnGlnTyrAsnAlaIleSerAlaThrThrTrpSerGlyGlnTyrVal 121
Db 314 GCTTGAAATPACTACCGCAACCGCACCTCCGACCAACGAGCGGCGGCGGAGTACGTC 373
Qy 122 GlyIleAlaGluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrGlnAla 141
Db 374 GCGGGCGCGGAGCGGAGTCAACACCCAGTGGCTGTGACTCCGCGCACACCGAGGCC 433
Qy 142 AsnAlaTrpIleSerThrLeuValGlyHisAspThrPheThrIleValIleProSerAla 161
Db 434 AACGCTGGAAATCCACGCTGCTCGGCCACGACCTTCAACCAAGTGAAGCGCTCGCC 493
Qy 162 AlaSerIleAspAlaAlaIleValIleValIleValIleValIleValIleValIleVal 181
Db 494 GCTTCATCGACGCGCGGAAGCGCGGCTCAACACGGAACCGGCTCGACCGCGTT 553
Qy 182 GlnGln 183
Db 554 CAGCAG 559
RESULT 9
AAQ53412
ID AAQ53412 standard; DNA: 552 BP.
XX
AC AAQ53412;
XX
DT 25-MAR-2003 (updated)
DT 27-JUN-1994 (first entry)
XX
DE Streptavidin gene.
XX
KM Streptavidin; protein secretion; Bacillus subtilis; ss.
XX
OS Streptomyces avidinii.
XX
FH Key Location/Qualifiers
FT CDS 1..552

```

FT      mat_peptide      73..519      a
FT      sig_peptide      1..72      b
FT      misc_feature      109..519      c
FT      /tag= d
FT      /note= "fused to npr signal peptide in pBE673"
XX
XX      W09324631-A1.
XX
XX      09-DEC-1993.
XX
XX      27-MAY-1993; 93WO-US05240.
XX
XX      29-MAY-1992; 92US-0891524.
XX
XX      (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX      Nagarajan V;
XX
XX      WPI, 1993-405822/50.
XX
XX      P-PSDB; AAR44491.
XX
XX      Streptavidin prodn. from Bacillus subtilis - using signal protein
XX      from bacterial exo-protein and expression element from Gram
XX      positive bacterial protein.
XX
XX      Disclosure; Fig 1b; 54pp; English.
XX
XX      Tetrameric biologically active streptavidin is produced by secretion
XX      from Bacillus subtilis transformed with a plasmid encoding the
XX      sequence.
XX      (updated on 25-MAR-2003 to correct PN field.)
XX
XX      SQ      Sequence 552 BP; 105 A; 207 C; 167 G; 70 T; 3 other;
XX
XX      Alignment Scores:
XX      Pred. No.:      3,12e-73      Length:      552
XX      Score:      927.00      Matches:      182
XX      Percent Similarity: 99.45%      Conservative: 0
XX      Best Local Similarity: 99.45%      Mismatches: 1
XX      Query Match: 99.04%      Indels: 0
XX      DB: 14      Gaps: 0
XX
XX      US-09-589-870B-2 (1-183) x AAO53412 (1-552)
XX
XX      QY      1 MetArgLysIleValValAlaAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
XX      1 ATGCGCAAGATCGTCGTGCGATCGCGCTTCCCTGACACAGGCTCGATTACGGCC 60
XX
XX      QY      21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGlnAgl 40
XX      61 AGGCGCTTCGCGAGATGCTCCAGAGACTCGAAGGCCCGGCTCGCGCCGACGCGCCG 120
XX
XX      QY      41 IleThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValThrAlaGlnValAsp 60
XX      121 ATCACCAGGACCTGATGACACCAAGCTCGGCTGACCTTATGATGACGCGGCGCCGAC 180
XX
XX      QY      61 GlyAlaLeuThrGlyThrTyrGlnSerAlaValGlyAsnAlaGlnSerArgTyrValLeu 80
XX      181 GCGCGCCCTGACCGCAACCTACGAGTCGCGCGGCAACGCCGACGCGCTACGCTCTG 240
XX
XX      QY      81 ThrGlyArgTyrAsnSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
XX      241 ACCGGTCGTTACGACAGCGCGCCCGGCGCACCGGCGGACCGGCGCCCTCGGTTGGAGC 300
XX
XX      QY      101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTyrSerGlyGlnTyr 120
XX      301 GTGGCTCGAAGATATACCTACCGCAACGCCCACTCGGAGCACCTGAGCGCGCCAGTAC 360
XX
XX      QY      121 ValGlyGlyAlaGlnAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGln 140

```

```

Db      361 GTCGCGCGCGCCGAGCGAGATCAACACCCAGTGCCTGTGACTCCGCGACACCGAG 420
QY      141 AlaAsnAlaTrpLysSerThrIleuValGlyHsAspThrPheThrLysValLysProSer 160
Db      421 GCCAAGCCTGGAAGTCCAGCGCTGCGCGCACGACCTTCACCAAGGTGAAGCGCTCC 480
QY      161 AlaAlaSerIleAspAlaAlaLysLysAlaGlyValAlaAsnGlyAsnProLeuAspAla 180
Db      481 GCGGCTTCATCGACCGCGCGGAGAAAGCGCGCTCAACACGCAACCGGCTCGAGCGC 540
QY      181 ValGlnGln 183
Db      541 GTTCAGCAG 549
XX
XX      RESULT 10
XX      AAV34715
XX      ID AAV34715 standard; DNA; 638 BP.
XX
XX      AC      AAV34715;
XX
XX      DT      27-AUG-1998 (first entry)
XX
XX      DE      S. avidinii streptavidin mutant #1 DNA.
XX
XX      KW      Streptavidin; ligand; binding affinity; mutant; isolation;
XX      purification; recover; immobilise; ss.
XX
XX      OS      Streptomyces avidinii.
XX
XX      FH      Key      Location/Qualifiers
XX      FT      CDS      50..601
XX      FT      sig_peptide 50..121
XX      FT      mat_peptide 122..598
XX      FT      /tag= c
XX      FT      /product= streptavidin
XX      FT      mutation 251..262
XX      FT      /tag= d
XX      FT      /note= "wild type GACTGCGCGCTC is replaced by
XX      GTCACGCGCGCT"
XX
XX      PN      EP835934-A2.
XX
XX      PD      15-APR-1998.
XX
XX      PE      09-OCT-1997; 97EP-0117504.
XX
XX      PR      10-OCT-1996; 96DE-1041876.
XX
XX      PA      (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
XX      PI      Skerra A, Voss S;
XX
XX      DR      WPI, 1998-218868/20.
XX      P-PSDB; AAW59217.
XX
XX      PT      Streptavidin mutants with higher binding affinity for peptide
XX      ligands - have mutation in amino acid region 44-53, used to isolate,
XX      purify or determine fusion proteins including these ligands
XX
XX      PS      Disclosure; Page -; 21pp; German.
XX
XX      SS      This sequence encodes a mutant streptavidin protein isolated from
XX      Streptomyces avidinii where the residues BSAV at position 44-47 of
XX      the mature wild type sequence are replaced by YTRK. This sequence is
XX      used to produce mutants which are used in a method to assay the binding
XX      affinity of streptavidin mutants. These mutants have a mutation within
XX      the amino acid (aa) region 44-53 of the wild-type protein show a higher
XX      binding affinity than the wild-type for peptide ligands that include
XX      the sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y
XX      and Z are both Gly, or Y = Gln and Z = Arg or Lys. Recombinant

```

streptavidin mutants can be used to isolate, purify and determine proteins or to determine/recover substances that contain streptavidin-binding groups. Such compounds may also be used to immobilise fusions on microtitre plates, microbeads or sensor chips. NOTE: This sequence does not appear in the specification but has been constructed from the wild-type streptavidin sequence represented in AAV34714.

SO Sequence 638 BP; 115 A; 244 C; 193 G; 86 T; 0 other;

Alignment Scores:

Pred. No.:	1,88e-72	Length:	638
Score:	919.00	Matches:	180
Percent Similarity:	98.91%	Conservative:	1
Best Local Similarity:	98.36%	Mismatches:	2
Query Match:	98.18%	Indels:	0
DB:	19	Gaps:	0

US-09-589-870B-2 (1-183) x AAV34715 (1-638)

```
QY 1 MetArgLysIleValAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
DB 50 ATGCGCAAGATCGTCTGACGACATCGCGTTCCCTGACCGATTCGATTACGGCC 109
QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGlnAlaGly 40
DB 110 AGCGCTTCGGCAGACCCCTCCAGACTCGAAGGCCAGCTTCGGCCCGCGAGCGCGC 169
QY 41 IleThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAlaAsp 60
DB 170 ATCACCAGGACCTGTCATCAACACGATCGGCTCAGCTTATCGTACCGGCGCGCGAC 229
QY 61 GlyAlaLeuThrGlyThrTyrGlySerAlaValGlyAlaAsnAlaGlnSerArgTyrValLeu 80
DB 230 GGGCGCCCTGACCGAAGCTTACGACGCGCGCGCGCGCGCGCGCGCTTACGCTCG 289
QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
DB 290 ACCGCTGCTTACACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCGATTGAGC 349
QY 101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyr 120
DB 350 GTGGCTGGAAAGATTAATCAACCGCAACGCCCACTCCGCGACCAAGTGAAGCCGCTCC 409
QY 121 ValGlyGlyValAlaGlnAlaArgIleAsnThrGlnTyrLeuLeuThrSerGlyThrThrGlu 140
DB 410 GTCGCGCGCGCGCGCGCGAGATCAACCGCAAGTGTGCTGCTCCGCGCACCGCGAG 469
QY 141 AlaAsnAlaTrpLysSerThrLysValGlyHisAspThrPheThrLysValLysProSer 160
DB 470 GCCAAGCGCTCGAAAGTCCACGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 529
QY 161 AlaAlaSerIleAspAlaAlaLysValGlyValAlaAsnAlaGlnProLeuAspAla 180
DB 530 GCCGCTTCATCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 589
QY 181 ValGlnGln 183
DB 590 GTTCACAGCAG 598
```

RESULT 11
AAV34716
ID AAV34716 standard; DNA; 638 BP.

XX AAV34716;
XX
AC AAV34716;
XX
DT 27-AUG-1998 (first entry)
XX
DE S. avidinii streptavidin mutant #2 DNA.
XX
KM Streptavidin; ligand; binding affinity; mutant; isolation;
XX purification; recover; immobilise; ss.

OS Streptomyces avidinii.
OS Synthetic.

Key	Location/Qualifiers
FT CDS	50..601
FT sig_peptide	/tag= a 50..121
FT mat_peptide	/tag= b 122..598
FT mutation	/tag= c /product= Streptavidin 251..262
FT	/tag= d /note= "Wild type GAGTCGCGCGC is replaced by ATCGTCGAGC"

EP835934-A2.

15-APR-1998.

09-OCT-1997; 97EP-0117504.

10-OCT-1996; 96DE-1041876.

(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

Skerra A, Voss S;

WPI; 1998-218868/20.

P-PSDB; AAW59218.

Streptavidin mutants with higher binding affinity for peptide ligands - have mutation in amino acid region 44-53, used to isolate, purify or determine fusion proteins including these ligands

Disclosure; Page -: 21pp; German.

This sequence encodes a mutant streptavidin protein isolated from Streptomyces avidinii where the residues E54V at position 44-47 of the mature wild type sequence are replaced by IGAR. This sequence is used to produce mutants which are used in a method to assay the binding affinity of streptavidin mutants. These mutants have a mutation within the amino acid (aa) region 44-53 of the wild-type protein show a higher binding affinity than the wild-type for peptide ligands that include the sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin mutants can be used to isolate, purify and determine proteins or to determine/recover substances that contain streptavidin-binding groups. Such compounds may also be used to immobilise fusions on microtitre plates, microbeads or sensor chips. NOTE: This sequence does not appear in the specification but has been constructed from the wild-type streptavidin sequence represented in AAV34714.

SO Sequence 638 BP; 116 A; 242 C; 194 G; 86 T; 0 other;

Alignment Scores:	2,83e-72	Length:	638
Pred. No.:	917.00	Matches:	180
Score:	917.00	Conservative:	0
Percent Similarity:	98.36%	Mismatches:	3
Best Local Similarity:	97.97%	Indels:	0
Query Match:	97.97%	Gaps:	0
DB:	19		

US-09-589-870B-2 (1-183) x AAV34716 (1-638)

```
QY 1 MetArgLysIleValAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
DB 50 ATGCGCAAGATCGTCTGACGACATCGCGTTCCCTGACCGATTCGATTACGGCC 109
QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGlnAlaGly 40
DB 110 AGCGCTTCGGCAGACCCCTCCAGACTCGAAGGCCAGCTTCGGCCCGCGAGCGCGC 169
```

Qy	41	Illehrhrglythrtpplyraaenqlhneuglyserthrphellvaltlrralaglyalaasp	60
Db	170	ATACCGGACCTGTGTACAAACCAAGCTCGCTCGACTTCATCTGTGACCGCGGGCGGAC	222
Qy	61	GIyAlaleuthrglyThrTYrGluSerAlaValGlyAsnAlagluSerArgTYrValleu	80
Db	230	GGCGCCTGAACCGGAACCTACATCGGTGCGAGGGGCAACGCCGAGACCGCTTAGCTCTG	289
Qy	81	ThhGlyArgTYrhapSerAlaProAlaThrAspGlySerGlyThrAlaleuGlyTYrThr	100
Db	290	ACCGTCGTATACGACACAGCGCCCGCCACGACGCGAGCGGACCGCCCTCGTTGACG	349
Qy	101	ValAlatrplyAsnAsnTYrArgAsnAlaHISerAlaThrTrpSerGlyGlnTYr	120
Db	350	GTCGGCTCGAGAGATTACTACCCGCAAGCCCACTCCGCGACCAAGTGAGCGGCACTAC	409
Qy	121	ValGlyGlyValagluAlaArgIleasnThrglnTrpLeuLeuThrSerGlyTYrThrglu	140
Db	410	GTCGGCGGCGCCGAGGAGGATCAACACCAGTGGTGTGACTTCGGCACCAACCGAG	469
Qy	141	AlaAsnAlaTrpIlysserThrIleuValGlyHisAspThrPheThrIlyValIysProSer	160
Db	470	GCCAAAGCCTGGAAGTCCACGCTGTCGGCCACGACCTTTCACCAAGGTGAAGCGTCC	529
Qy	161	AlaAlaSerIleAspAlaAlaIlySlySAlaGlyValAsnAsnGlyAsnProIleAspAla	180
Db	530	GCGGCTTCATCGACCGCGGAGAAAGCGCGCTCAACACGCAACCGCTCGACGCC	589
Qy	181	ValGlnGln 183	
Db	590	GTTTCAGCAG 598	
RESULT 12			
ID	AAC86563 standard; DNA; 1239 BP.		
XX	AAC86563;		
AC	AAC86563;		
DT	02-APR-2001 (first entry)		
XX			
DE	DNA encoding a fusion of anti-CD20 single chain antibody/streptavidin.		
XX			
KW	Streptavidin; tumour cell; cancer; adenocarcinoma;		
KW	hematological malignancy; ss.		
XX			
OS	Synthetic.		
OS	Streptomyces avidinii.		
OS	Homo sapiens.		
XX			
PN	WO200075333-A1.		
XX			
PD	14-DEC-2000.		
XX			
PF	05-JUN-2000; 2000WO-US15595.		
XX			
PR	07-JUN-1999; 99US-0137900.		
PR	03-DEC-1999; 99US-0168976.		
XX			
PA	(NEOR-) NEORX CORP.		
XX			
PI	Goshorn SC, Graves SS, Schultz JE, Lin Y, Sanderson JA, Reno JM;		
XX			
DR	MPI; 2001-091213/10.		
XX			
DR	P-PSDB; AAB30694.		
XX			
PT	New vector constructs for expressing genomic streptavidin fusion		
XX	proteins which are useful for targeting tumour cells associated with		
XX	cancer. e.g. adenocarcinomas -		
XX			
PS	Example 2; Fig 11A; 100pp; English.		
XX			
CC	The present sequence encodes a fusion of an anti-CD20 single chain		

antibody (B959) streptavidin. The fusion protein is expressed using vectors of the invention. The specification describes vector constructs for expressing streptavidin fusion proteins. The vector comprises a nucleic acid encoding streptavidin or its functional variant operatively linked to a promoter, and a cloning site for insertion of a second nucleic acid sequence encoding a polypeptide to be fused with streptavidin. Interposed between the promoter and the first nucleic acid sequence. Alternatively, the vector construct comprises a nucleic acid, operatively linked to a promoter, encoding a polypeptide to be fused with streptavidin, and a cloning site for insertion of a second nucleic acid encoding at least 129 amino acids of streptavidin or its functional variant. The fusion proteins are useful for targeting tumour cells, particularly tumour cells associated with cancer. e.g. adenocarcinomas or hematological malignancies. The vector construct is useful for expressing of streptavidin fusion proteins. In particular, these are useful as tools for medical diagnostics and therapeutic purposes, e.g. for detecting the presence or absence of, or treating, a target site within a mammalian host.

Sequence 1239 BP; 270 A; 392 C; 356 G; 221 T; 0 other;

Alignment Scores:

Score: 855.50

Best Local Similarity:

DB: 22

US-09-589-870B-2 (1-183) X AAC86563 (1-1239)

QY 14 ThrThrValSerIleThrAla---SerAlaSerAlaAspProSerIleLysAspSerLysAla 32

Db 724 ACCACGGTCACCGTGAGCTCTGGCTCTGGTTCGGCAGACCCCTCCAAGGACTCGAAGGCC 78

QY 33 GlnValSerAlaAlaGluAlaGlyIleThrGlyThrTrpTyrAsnGlnLeuGlySerThr 52

Db 784 CAGGTCTCGCCGCCGAGGCCGCATCACCGGCACCTGTACTACAACAGCTCGGCTCGACC 84

QY 53 PheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyrGluSerAlaValGly 72

Db 844 TTCATCGTACCGCGCGCCCTGACCGGAACCTACGAGTCGGCCGTGGC 90

73 AsnAlaGluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAspGly 92

904 AACGCCGAGAGCCGCTACGTCCTGACCGGTCGTTACGACAGCGCCCCGGCCACCGACGGC 96

93 serGlylnraAlaLeuGlyTrrpThrValAlaTrrpLysAsnAsnTYrArgAsnAlaHisSer 11

964 AGCGCACCGCCCTCGGTGGACGGTGGCCCTGGAGAATACTACCGCAAGCCCACTCC 10

113 M A I N I I N I T P S E R G I Y G I N I Y V A L G I Y G I Y A L A G I U A L A R G I L E A S N T H R G I N T R P 13

1024 GCGACACGAGAGCGCCAGIACGICGGCGGCGCCGAGCGAGGATCAACACCCAGTGG 10

133 Lebeuimsergi inirigulaasthalairplysser inrleuValGlyh1sAsp 15

2081 CAGCAGCCACGCGGGTCTGTTGGGTCACGAC TT

409 408 407 406 405 404 403 402 401 400 399 398 397 396 395 394 393 392 391 390 389 388 387 386 385 384 383 382 381 380 379 378 377 376 375 374 373 372 371 370 369 368 367 366 365 364 363 362 361 360 359 358 357 356 355 354 353 352 351 350 349 348 347 346 345 344 343 342 341 340 339 338 337 336 335 334 333 332 331 330 329 328 327 326 325 324 323 322 321 320 319 318 317 316 315 314 313 312 311 310 309 308 307 306 305 304 303 302 301 300 299 298 297 296 295 294 293 292 291 290 289 288 287 286 285 284 283 282 281 280 279 278 277 276 275 274 273 272 271 270 269 268 267 266 265 264 263 262 261 260 259 258 257 256 255 254 253 252 251 250 249 248 247 246 245 244 243 242 241 240 239 238 237 236 235 234 233 232 231 230 229 228 227 226 225 224 223 222 221 220 219 218 217 216 215 214 213 212 211 210 209 208 207 206 205 204 203 202 201 200 199 198 197 196 195 194 193 192 191 190 189 188 187 186 185 184 183 182 181 180 179 178 177 176 175 174 173 172 171 170 169 168 167 166 165 164 163 162 161 160 159 158 157 156 155 154 153 152 151 150 149 148 147 146 145 144 143 142 141 140 139 138 137 136 135 134 133 132 131 130 129 128 127 126 125 124 123 122 121 120 119 118 117 116 115 114 113 112 111 110 109 108 107 106 105 104 103 102 101 100 99 98 97 96 95 94 93 92 91 90 89 88 87 86 85 84 83 82 81 80 79 78 77 76 75 74 73 72 71 70 69 68 67 66 65 64 63 62 61 60 59 58 57 56 55 54 53 52 51 50 49 48 47 46 45 44 43 42 41 40 39 38 37 36 35 34 33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

22 12345678910111213141516171819202122232425262728293031323334353637383940414243444546474849505152535455565758596061626364656667686970717273747576777879808182838485868788899091929394959697989910010110210310410510610710810911011111211311411511611711811912012112212312412512612712812913013113213313413513613713813914014114214314414514614714814915015115215315415515615715815916016116216316416516616716816917017117217317417517617717817918018118218318418518618718818919019119219319419519619719819920020120220320420520620720820921021121221321421521621721821922022122222322422522622722822923023123223323423523623723823924024124224324424524624724824925025125225325425525625725825926026126226326426526626726826927027127227327427527627727827928028128228328428528628728828929029129229329429529629729829930030130230330430530630730830931031131231331431531631731831932032132232332432532632732832933033133233333433533633733833934034134234334434534634734834935035135235335435535635735835936036136236336436536636736836937037137237337437537637737837938038138238338438538638738838939039139239339439539639739839940040140240340440540640740840941041141241341441541641741841942042142242342442542642742842943043143243343443543643743843944044144244344444544644744844945045145245345445545645745845946046146246346446546646746846947047147247347447547647747847948048148248348448548648748848949049149249349449549649749849950050150250350450550650750850951051151251351451551651751851952052152252352452552652752852953053153253353453553653753853954054154254354454554654754854955055155255355455555655755855956056156256356456556656756856957057157257357457557657757857958058158258358458558658758858959059159259359459559659759859960060160260360460560660760860961061161261361461561661761861962062162262362462562662762862963063163263363463563663763863964064164264364464564664764864965065165265365465565665765865966066166266366466566666766866967067167267367467567667767867968068168268368468568668768868969069169269369469569669769869970070170270370470570670770870971071171271371471571671771871972072172272372472572672772872973073173273373473573673773873974074174274374474574674774874975075175275375475575675775875976076176276376476576676776876977077177277377477577677777877978078178278378478578678778878979079179279379479579679779879980080180280380480580680780880981081181281381481581681781881982082182282382482582682782882983083183283383483583683783883984084184284384484584684784884985085185285385485585685785885986086186286386486586686786886987087187287387487587687787887988088188288388488588688788888989089189289389489589689789889990090190290390490590690790890991091191291391491591691791891992092192292392492592692792892993093193293393493593693793893994094194294394494594694794894995095195295395495595695795895996096196296396496596696796896997097197297397497597697797897998098198298398498598698798898999099199299399499599699799899910001001100210031004100510061007100810091010101110121013101410151016101710181019102010211022102310241025102610271028102910301031103210331034103510361037103810391040104110421043104410451046104710481049105010511052105310541055105610571058105910601061106210631064106510661067106810691070107110721073107410751076107710781079108010811082108310841085108610871088108910901091109210931094109510961097109810991100110111021103110411051106110711081109111011111112111311141115111611171118111911201121112211231124112511261127112811291130113111321133113411351136113711381139114011411142114311441145114611471148114911501151115211531154115511561157115811591160116111621163116411651166116711681169117011711172117311741175117611771178117911801181118211831184118511861187118811891190119111921193119411951196119711981199120012011202120312041205120612071208120912101211121212131214121512161217121812191220122112221223122412251226122712281229123012311232123312341235123612371238123912401241124212431244124512461247124812491250125112521253125412551256125712581259126012611262126312641265126612671268126912701271127212731274127512761277127812791280128112821283128412851286128712881289129012911292129312941295129612971298129913

103
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846

[illegible]

AAC86562

XX
XX

1.
2.
3.
4.
5.
6.
7.
8.
9.
10.
11.
12.
13.
14.
15.
16.
17.
18.
19.
20.
21.
22.
23.
24.
25.
26.
27.
28.
29.
30.
31.
32.
33.
34.
35.
36.
37.
38.
39.
40.
41.
42.
43.
44.
45.
46.
47.
48.
49.
50.

XX


```

DT 02-APR-2001 (first entry)
DE DNA encoding a fusion of single chain antibody/streptavidin.
XX
XX Streptavidin; tumour cell; cancer; adenocarcinoma;
XX hematological malignancy; huNR-LU-10; EGP40; EPCAM; ss.
XX
XX Synthetic.
OS Streptomyces avidinii.
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX CDS 311..1606
XX FT /tag= a
XX FT /product= "huNR-LU-10 and streptavidin fusion"
XX
XX MO200075333-A1.
XX
XX 14-DEC-2000.
XX
XX 05-JUN-2000; 2000MO-US15595.
XX
XX 07-JUN-1999; 99US-0137900.
XX PR 03-DEC-1999; 99US-0168976.
XX
XX (NEOR-) NEORX CORP.
XX
XX Goshorn SC, Graves SS, Schultz JE, Lin Y, Sanderson JA, Reno JW,
XX WPI: 2001-091213/10.
XX DR P-PSDB; AAB30693.
XX
XX New vector constructs for expressing genomic streptavidin fusion
XX proteins which are useful for targeting tumour cells associated with
XX cancer, e.g. adenocarcinomas -
XX
XX Example 1; Fig 10; 100pp; English.
XX
XX The present sequence encodes a fusion of the single chain antibody
XX huNR-LU-10 and streptavidin. The antibody binds the antigen EGP40 or
XX EPCAM. The fusion protein is expressed using vectors of the invention.
XX
XX The specification describes vector constructs for expressing streptavidin
XX fusion proteins. The vector comprises a first nucleic acid encoding
XX streptavidin or its functional variant operatively linked to a promoter,
XX and a cloning site for insertion of a second nucleic acid sequence
XX encoding a polypeptide to be fused with streptavidin, intersposed between
XX the promoter and the first nucleic acid sequence. Alternatively, the
XX vector construct comprises a first nucleic acid, operatively linked to a
XX promoter, encoding a polypeptide to be fused with streptavidin, and a
XX cloning site for insertion of a second nucleic acid encoding at least
XX 129 amino acids of streptavidin or its functional variant. The fusion
XX proteins are useful for targeting tumour cells, particularly tumour cells
XX associated with cancer, e.g. adenocarcinomas or hematological
XX malignancies. The vector construct is useful for expressing of
XX streptavidin fusion proteins. In particular, these are useful as tools
XX for medical diagnosis and therapeutic purposes, e.g. for detecting the
XX presence or absence of, or treating, a target site within a mammalian
XX host.
XX
XX Sequence 1612 BP; 353 A; 493 C; 459 G; 307 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 2,8e-66 Length: 1612
XX Score: 854.50 Matches: 166
XX Percent Similarity: 97.67% Conservative: 2
XX Best Local Similarity: 96.51% Mismatches: 3
XX Query Match: 91.29% Indels: 1
XX Gaps: 1
XX
XX US-09-589-870B-2 (1-183) x AAC66562 (1-1612)

```

Dd		1091	ACCTTAAGTCAACCGGAGC---TCGGCTCTGGTTCCGACAACCCTTCGAAGACTCGAAG	1147
Oy		32	AAGInValSerAlaIaGuaIaGlyIleThrGlyThrTrpTyraSngInLeuGlySer	51
Dd		1148	GCCCAAGGTCTCGGCCGCCGAGGCCGGCATCACCGGCACCTGGTAACAACGAGCTCGGCTCG	1207
Oy		52	ThrTheileValIThrAlaGlyAlaAPRGlyAlaLeuTherGlyThrTyrgIuseRAlaVal	71
Dd		1208	ACCTTCATCGTGAGCCCGGGCGCGGACGGGGCCCTTAGCCGGAACCTAACGATCGGCCCTC	1267
Oy		72	GlyAmnIaIuseARgTYrVallLeuThrGlyYArGYrTyraSPserAlaProAlaThrAsp	91
Dd		1268	GCGAACCGCCAGACCCCCCTACGTCCTTAACCGGTCGTAAACAGCCGCCCGCCACCGAG	1327
Oy		92	GlySerGIyThralaLeuGIyTrpThrValAlaTrplysaAnanTYrArgAnaAlaHis	111
Dd		1328	GGCAGCGGCACCGCCCTCGGTTGGACGGTGGCTCGAAGAATAACTACCGCAACGCCAC	1387
Oy		112	SerAlaThrThrTrpSerGIyGlnTYrValIGlyGlyAlaGluAlaArgIIeaNthrGln	131
Dd		1388	TCCCCGACCAAGCTGAGACCGGCCAGTACGTGGGGCGGCCGAGCGAGATCAACACCAC	1447
Oy		132	TrpleuueuthrseryIyThrrThruAlaAsnaAlaTrplyseSerThrlauValIGlyHis	151
Dd		1448	TGGCTGCTGACCTTCGGCACCAACGAGCCAGCCCTCGAAGTCCACGCTGTGTGGCCAC	1507
Oy		152	AspThrPheThrLyysValLysProSerAlaAlaSerIIeasPAIaAlaLysAlaGly	171
Dd		1508	GACACCTTCACCAAGGTGAAGCCGTCGGCGGCTTCATCCAACCGCGGAAGAGCGCGGC	1567
Oy		172	ValAsnaSngIyaenProLeuAspaAlaValGlnGln	183
Dd		1568	GTCACAACAGCGCAACCCGCTCGACGCGCTTCACAGC	1603
RESULT 14				
ID	AAC86564		standard; DNA; 1280 BP.	
XX	XX	AC	AAC86564;	
XX	DT	02-APR-2001	(first entry)	
DE	XX	Anti-CD20 single chain antibody/streptavidin fusion protein cassette.		
KW	XX	Streptavidin; tumour cell; cancer; adenocarcinoma;		
KW	XX	hematological malignancy; ss.		
OS	XX	Synthetic.		
OS	XX	Strepcomycas avidinii.		
OS	XX	Homo sapiens.		
FT	FH	Key	Location/Qualifiers	
FT	CDS	3..1274	/cag= a	
FT		/product= "anti-CD20 scFv and streptavidin fusion"		
Pt	PD	WO200075333-A1.		
XX	PD	14-DEC-2000.		
PF	PP	05-JUN-2000; 2000MO-US15595.		
XX	PR	07-JUN-1999; 99US-0137900.		
XX	PR	03-DEC-1999; 99US-0168976.		
XX	PA	(NEOR-) NEORX CORP.		
XX	PI	Goshorn SC, Graves SS, Schultz JE, Lin Y, Sanderson JA, Reno JM;		
XX	DR	WI: 2001-091213/10.		
XX	DR	P-PSDB; AABJ30695.		
PT	XX	New vector constructs for expressing genomic streptavidin fusion		

PT proteins which are useful for targeting tumour cells associated with
 PT cancer, e.g. adenocarcinomas -

XX Example 2; Fig 11C; 100bp; English.

CC The present sequence encodes a fusion of an anti-CD20 single chain
 CC antibody (B99) streptavidin. The fusion protein is expressed using
 CC vectors of the invention. The specification describes vector constructs
 CC for expressing streptavidin fusion proteins. The vector comprises a
 CC nucleic acid encoding streptavidin or its functional variant operatively
 CC linked to a promoter, and a cloning site for insertion of a second
 CC nucleic acid sequence encoding a polypeptide to be fused with
 CC streptavidin, interposed between the promoter and the first nucleic
 CC acid sequence. Alternatively, the vector construct comprises a nucleic
 CC acid, operatively linked to a promoter, encoding a polypeptide to be
 CC fused with streptavidin, and a cloning site for insertion of a second
 CC nucleic acid encoding at least 129 amino acids of streptavidin or its
 CC functional variant. The fusion proteins are useful for targeting tumour
 CC cells, particularly tumour cells associated with cancer.
 CC e.g. adenocarcinomas or hematological malignancies. The vector construct
 CC is useful for expressing of streptavidin fusion proteins. In particular,
 CC these are useful as tools for medical diagnostics and therapeutic
 CC purposes, e.g. for detecting the presence or absence of, or treating, a
 CC target site within a mammalian host.

SO Sequence 1280 BP; 267 A; 397 C; 388 G; 228 T; 0 other;

Alignment Scores:

Pred. No.: 9, 81e-66 Length: 1280
 Score: 847.00 Matches: 162
 Percent Similarity: 98.79% Conservative: 1
 Best Local Similarity: 98.18% Mismatches: 2
 Query Match: 90.49% Indels: 0
 DB: 22 Gaps: 0

US-09-589-870B-2 (1-183) x AAC86564 (1-1280)

QY 19 ThAlaSerAlaSerAlaSPProSerLyASpSerLyAlaGlnValSerAlaAlaGln 38
 DB 777 TCTGGCTCTGTTCTGGCAGACCCCTCAAGACTCCAGAGGCTCTGGCCGCGAG 836
 QY 39 AlaGlyIleThrGlyThrTrpTyraSngInLeuGlySerThPheIleValhrAlaGly 58
 DB 837 GCCGGATACCGGCACCTGTCACACAGCTGCTGCTCACTTCATCTGACCCGCGGC 896
 QY 59 AlaSPGlyAlaLeuThrGlyThrTyrgInuSerAlaValGlyASnaIaGlnuSerArgTy 78
 DB 897 GCCGAGCGCGCCTGACCGGAACCTAGAGTCGGCGCTGGCAAGCCGAGCGGCTAC 956
 QY 79 ValIleThrGlyArgTyraSPSerAlaProAlaThrASpGlySerGlyThrAlaLeuGly 98
 DB 957 GTCTCGAACCGGTGTTACACAGCGCCCGCCACCGAGGAGGCGCACCCCTCCGT 1016
 QY 99 TrpThrValAlaTrpTyraSnaSntTyraGnaAlaAhIaSerAlaThrThrTrpSerGly 118
 DB 1017 TGGACGGTGGCTGGAGAAGATAACTACCGCAACGCCCACTCCGCAACACGTGGAGCGGC 1076
 QY 119 GlnTyraValGlyValAlaGlnAlaArgIleAsnThrGlnTrpLeuThrSerGlyThr 138
 DB 1077 CAGTACGTGGCGGCGCCGAGCGAGATCAACACCCAGTGGCTGCTGACCTCCGCGACC 1136
 QY 139 ThrGlnAlaASnaIaTrpTyraSerThrLeuValGlyHISepThrPheThrLyValys 158
 DB 1137 ACCGAGGCCAACGCTGGAGATCCAGCGTGGCGCCAGACACTTCAACCAAGGTGAG 1196
 QY 159 ProSerAlaAlaSerIleASpAlaAlaLySlyAlaGlyValaASnaGlyASnProLeu 178
 DB 1197 CGGTCCGCGCGCTTCATGACCGCGCGGAAGAGCGGCTCAACACGCAACCCGCTC 1256
 QY 179 AspAlaValGlnGln 183
 DB 1257 GACGCCGTTACAGAG 1271

RESULT 15
 ID AAQ70659 standard; DNA; 1356 BP.

XX AC AAQ70659;
 XX AC
 DT 25-MAR-2003 (updated)
 DT 26-MAR-1995 (first entry)
 XX XX
 DE ScfV pRAS108 and pRAS112 DNA.

XX Amplification; single chain variable region fusion protein; PCR; ss.

OS Synthetic.

FH Key Location/Qualifiers
 FT CDS 40..1348
 FT /*tag= a

PN WO9415644-A1.

XX 21-JUL-1994.

PF 17-JAN-1994; . 94WO-GB00087.

PR 15-JAN-1993; . 93GB-0000686.

PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.

XX Deonaraia M, Epenetos AA, Spooner RA;

XX WPI; 1994-248907/30.

DR P-PSDB; AAR56483.

XX New cpds. comprising a targeting portion and a cytotoxic portion
 PT - used esp. for treating mammals for destroying target cells,
 PT partic. tumour cells

XX Claim 36; Fig 9; 114pp; English.

CC The sequence is that of the ScfV pRAS108 and pRAS112 between HindIII and
 CC EcoRI site obtd. by PCR.
 CC See also AAQ70652-69.
 CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 1356 BP; 332 A; 359 C; 359 G; 306 T; 0 other;

Alignment Scores:

Pred. No.: 1, 29e-65 Length: 1356
 Score: 846.00 Matches: 162
 Percent Similarity: 99.39% Conservative: 1
 Best Local Similarity: 98.78% Mismatches: 1
 Query Match: 90.38% Indels: 0
 DB: 15 Gaps: 0

US-09-589-870B-2 (1-183) x AAQ70659 (1-1356)

QY 20 AlaSerAlaSerAlaSPProSerLyASpSerLyAlaGlnValSerAlaAlaGln 39
 DB 853 GCTGCCACCTGACGACCGCTCCAGAGACTCCAAAGCTTACGTTTCTGAGCGGAAGCT 912
 QY 40 GlyIleThrGlyThrTrpTyraSngInLeuGlySerThPheIleValhrAlaGlyAla 59
 DB 913 GGTATCACTGGACCTGCTGATTAACCACTGGGCTGCTTATTTGATCCCTGGTGGC 972
 QY 60 AspGlyAlaLeuThrGlyThrTyrgInuSerAlaValGlyASnaIaGlnuSerArgTy 79
 DB 973 GACGAGCTTGTGCTGACCTAGATCGAGATCTGCGTTGGTAACCAAGATCCCTACGTA 1032
 QY 80 LeuThrGlyArgTyraSPSerAlaProAlaThrASpGlySerGlyThrAlaLeuGlyTrp 99
 DB 1033 CTGACTGGCCGTTATAGCTGCACTGCCACCGAGATGCTCTGTGTAACCCCTGCGGCTGG 1092

```
QY 100 ThrValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGln 119
Db 1093 ACTGTGGCTTGGAAAAACAACATCGTAATGCCACAGCCCACTAGTGTCTGGCCAA 1152
QY 120 TyrValGlyGlyValArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThr 139
Db 1153 TACGTTGGCGGCTGCTGAGCTCGTATCAACACTCAGTGGCTGTTAACATCGGCACTACC 1212
QY 140 GluAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValLysPro 159
Db 1213 GAAGCGAATGATGAAATGACACTAGTAGTCAATGACACCTTTACCAAGTTAAGCCT 1272
QY 160 SerAlaAlaSerIleAspAlaAlaLysValGlyValAsnGlyAsnProLeuAsp 179
Db 1273 TCTGCTGCTAGCATTTGATGCTGCCAAGAAAGCAGGCGCTAAACAGGTAACTCTTAGAC 1332
QY 180 AlaValGlnGln 183
Db 1333 GCTGTTACGCAA 1344
```

Search completed: October 28, 2003, 02:09:54
Job time : 257 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 27, 2003, 21:54:43 ; Search time 3292 Seconds

(without alignments)
2274.138 Million cell updates/sec

Title: US-09-589-870B-2

Sequence: 1 MKKIVAAIAVSLTVSIFA.....IDAKKAGVNGNPLDAVQ 183

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame_p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.epool/US09589870/runac_27102003.104445.16617/app_query.fasta_1.327
-DB=GenEmbl -QMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCL
-OUTFMT=pro -NORM=ext -HEADING=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09589870@CCN_1_1_3508@runac_27102003.104445.16617 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:.*
1: gb_da.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_scs.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_scs.*
28: em_un.*

29: em_vi.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rod.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_sy.*
39: em_hcgo_hum.*
40: em_hcgo_mus.*
41: em_hcgo_other.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	936	100.0	638	1	SASTRAVG	X03591 Streptomyces
2	936	100.0	638	6	A93649	A93649 Sequence 1
3	936	100.0	638	6	AR177252	AR177252 Sequence
4	936	100.0	638	6	AR204862	AR204862 Sequence
5	936	100.0	638	6	AR217936	AR217936 Sequence
6	936	100.0	638	6	AX057941	AX057941 Sequence
7	933	99.7	638	6	I01349	I01349 Sequence 2
8	932	99.6	625	1	S78777	S78777 streptavidin
9	897	95.8	625	1	S78782	S78782 streptavidin
10	855.5	91.4	1239	6	AX057945	AX057945 Sequence
11	854.5	91.3	1612	6	AX057943	AX057943 Sequence
12	847	90.5	1280	6	AX057947	AX057947 Sequence
13	846	90.4	1356	6	A93965	A93965 Sequence 4
14	846	90.4	1356	6	A93152	A93152 Sequence 4
15	846	90.4	1356	6	AR082490	AR082490 Sequence
16	837.5	88.5	525	6	A20698	A20698 Fxa-StrpA D
17	837.5	88.5	525	6	I15647	I15647 Sequence 2
18	835	89.2	498	6	I09059	I09059 Sequence 38
19	834	89.1	483	12	AF283893	AF283893 Synthetic
20	834	89.1	498	6	A93838	A93838 Sequence 7
21	834	89.1	507	6	A00743	A00743 S.avidinii
22	834	89.1	1266	6	AX473375	AX473375 Sequence
23	834	89.1	1266	6	BD144212	BD144212 Fused pro
24	772	82.5	1881	6	A93967	A93967 Sequence 6
25	745	79.6	1296	6	A93154	A93154 Sequence 6
26	745	79.6	1296	6	AR082491	AR082491 Sequence
27	745	79.6	1296	6	AR231457	AR231457 Sequence
28	687.5	73.5	1173	6	BD140614	BD140614 Bifunctional
29	686	72.9	1176	6	AR231456	AR231456 Sequence
30	682	72.9	4354	12	ASV14583	Y14583 Artificial
31	682	72.9	4354	12	SCFV18290	Y18290 Expression
32	682	72.9	1257	6	A93569	A93569 Sequence 8
33	681.5	72.8	1257	6	A93156	A93156 Sequence 8
34	681.5	72.8	1257	6	AR082492	AR082492 Sequence
35	680	72.6	399	12	AGSTREP	X65082 Synthetic
36	679	72.5	384	6	AX348055	AX348055 Sequence
37	679	72.5	384	6	AX353922	AX353922 Sequence
38	679	72.5	1254	12	AR146718	AR146718 Synthetic
39	679	72.5	2019	6	E10169	E10169 DNA encodin
40	679	72.5	384	6	A93663	A93663 Sequence 15
41	675	72.1	384	6	AR177264	AR177264 Sequence
42	675	72.1	384	6	AR217948	AR217948 Sequence
43	675	72.1	387	6	I17472	I17472 Sequence 1
44	675	72.1	387	6	E10170	E10170 DNA encodin
45	660	70.5	2055	6	E10170	E10170 DNA encodin

ALIGNMENTS

RESULT 1

SASTRAVG
LOCUS 638 bp DNA linear BCT 30-MAR-1995
DEFINITION Streptomyces avidinii gene for streptavidin.
ACCESSION X03591
VERSION X03591.1 GI:46740
KEYWORDS streptavidin.
SOURCE Streptomyces avidinii
ORGANISM Streptomyces avidinii
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 638)
Argarana, C.E., Kuntz, I.D., Birken, S., Axel, R. and Cantor, C.R.
Molecular cloning and nucleotide sequence of the streptavidin gene
Nucleic Acids Res. 14 (4), 1871-1882 (1986)
JOURNAL 86148514
MEDLINE 3951999
PUBMED
COMMENT Data kindly reviewed (13-JUL-1986) by C. Argarana.
FEATURES
source
1. 638
Location/Qualifiers
/organism="Streptomyces avidinii"
/mol_type="genomic DNA"
/db_xref="taxon:1895"
50..601
/note="put. streptavidin precursor"
/codon_start=1
/transl_table=1
/protein_id="CAA27265.1"
/db_xref="GI:46741"
/translation="MKRIYVAIAVSLTIVSTASADPSKDSKQVSAARAGITGT
WYNQGSTIVTAGDAGLTGYESVGAESRYVLGKRYDSAPATGSGTALGMWIVA
WNNYRNHSAITWSGOYVGAEARINTQWLTSGTTEANMKSITLVGHDTFTKVPKPS
AASIDAKKAGVNNGNPLDAVQ"
50..121
/note="put. signal peptide (aa -24 to -1)"
122..598
/product="mat. streptavidin (aa 1-160)"
CDS
115 a 244 c 193 g 86 t
ORIGIN
BASE COUNT 115 a 244 c 193 g 86 t
ALIGNMENT SCORES
Pred. No.: 3 24e-63 Length: 638
Score: 936.00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 1
US-09-589-870B-2 (1-183) x SASTRAVG (1-638)
QY 1 MetArgLysIleValAlaAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
DB 50 ATGCGCAAGATCGTTCGTCAGCCATCGCGCTTCCCTGACACAGCGTCTCGATTACGGCC 109
QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaAlaGly 40
DB 110 AGCGCTTCGGAGAGCCCTCCCAAGACTCGAAGGCCCGAGTCTCGCGCCGAGGCCGCGC 169
QY 41 IleThrGlyThrTyrTrpAsnGlnLeuGlySerThrPheIleValThrAlaGlyAlaAsp 60
DB 170 ATCACCAGGACCTGTTACACAGCTCGGCTCGACCTTCATCTGTAACCGGAGCCGCAC 229
QY 61 GialaLeuThrGlyThrTyrGlnSerAlaValGlyAsnAlaGlySerArgTyrValLeu 80
DB 230 GCGCGCCGACCGGACCTAGCAGAGTCGCGCGTCCGCAAGCGGAGCGGCTAGCTCTCG 289
QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
DB 290 ACCGGTCTTACGAGAGCGCCCGGCGCACCGGAGCGGAGCGGCGCTCGGTTGAGCG 349
QY 101 ValAlaTPrpLysAsnAspTyrArgAsnAlaHisSerAlaThrThrTyrSerGlyGlnTyr 120
DB 350 GTGGCTCGAGAAATACCTACCGCAAGCCCACTCCGCGACCAAGTGAAGCGGCCAGTAC 409

QY 121 ValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuThrSerGlyThrThrGly 140
DB 410 GTGGCGGGCGCGGAGCGGAGATCAACCCAGTGGCTGCTGACCTCGGACACCGAG 469
QY 141 AlaAsnAlaTPrpLysSerThrLeuValGlyHisAspThrPheThrIleValLysProSer 160
DB 470 GCCAAGCGCTGGAAGTCCACGCTGTCGCCACGACACCTTCACCAAGTGAAGCCGTC 529
QY 161 AlaAlaSerIleAspAlaAlaLysAlaGlyValaAsnGlnGlyAsnProLeuAspAla 180
DB 530 GCGGCTCGATCGACGCGGCAAGAGCGCGGCTCAACCAAGCGACCCCTCGACGCC 589
QY 181 ValGlnGln 183
DB 590 GTTCAGCAG 598
RESULT 2
A93649 638 bp DNA linear PAT 22-JAN-2000
LOCUS A93649
DEFINITION Sequence 1 from Patent EP0799890.
ACCESSION A93649
VERSION A93649.1 GI:6741838
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 638)
Mueller, R.D. and Deger, A.D.
Recombinant inactive core streptavidin mutants
Patent: EP 0799890-A 1 08-OCT-1997;
JOURNAL BOEHRINGER MANNHEIM GMBH (DE)
FEATURES
source
1. 638
Location/Qualifiers
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
50..601
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB69587.1"
/db_xref="GI:6741838"
/translation="MKRIYVAIAVSLTIVSTASADPSKDSKQVSAARAGITGT
WYNQGSTIVTAGDAGLTGYESVGAESRYVLGKRYDSAPATGSGTALGMWIVA
WNNYRNHSAITWSGOYVGAEARINTQWLTSGTTEANMKSITLVGHDTFTKVPKPS
AASIDAKKAGVNNGNPLDAVQ"
50..121
/note="put. signal peptide (aa -24 to -1)"
122..598
/product="mat. streptavidin (aa 1-160)"
CDS
115 a 244 c 193 g 86 t
ORIGIN
BASE COUNT 115 a 244 c 193 g 86 t
ALIGNMENT SCORES
Pred. No.: 3 24e-63 Length: 638
Score: 936.00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
DB: 6
US-09-589-870B-2 (1-183) x A93649 (1-638)
QY 1 MetArgLysIleValAlaAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
DB 50 ATGCGCAAGATCGTTCGTCAGCCATCGCGCTTCCCTGACACAGCGTCTCGATTACGGCC 109
QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaAlaGly 40
DB 110 AGCGCTTCGGAGAGCCCTCCCAAGACTCGAAGGCCCGAGTCTCGCGCCGAGGCCGCGC 169
QY 41 IleThrGlyThrTyrTrpAsnGlnLeuGlySerThrPheIleValThrAlaGlyAlaAsp 60
DB 170 ATCACCAGGACCTGTTACACAGCTCGGCTCGACCTTCATCTGTAACCGGAGCCGCAC 229

QY 61 G1A1A1euthrG1yThrTg1uSerAlaValG1yAenAlaG1uSerArG1yVal1leu 80
|
|
|
Db 230 GGGGCCCTGACCGGAACCTACAGAGTCGGCCGTCCGCAACCGCAGACCCCTACGTCCTG 289
|
|
|
QY 81 ThrG1yArG1yThrAspSerAlaProAlaThrAspG1ySerG1yThra1A1eug1yTTPThr 100
|
|
|
Db 290 ACCGGTCGTACGACAGCCGCCGCCAGCAGCGGCGGACCCGCTCGGTGGAGC 349
|
|
|
QY 101 ValAla1eThrLyAspAenTyrArgAenAlaHisSerAlaThrThrTPSerG1yG1nTyr 120
|
|
|
Db 350 GTGGCCCTGGAAGATACACCGCAAGCCCACTCCCGCAGCAGCTGGAGCCGCGATAC 409
|
|
|
QY 121 ValG1yG1yAlaG1uAlaArg1leAenThrG1nTTPLeuThrSerG1yThrThrG1u 140
|
|
|
Db 410 GTCCGGCGGCCCGCAGCGAGATCAACACCACTGGCTGCTACCTCCGCGACACCGAG 469
|
|
|
QY 141 AlaAenAla1eThrLySerThrLeuValG1yHisAspThrPheThrLyVal1yProSer 160
|
|
|
Db 470 GCCAAGCCTGGAAGTCACAGCTGGTGGCCACGACCTTCAACAAAGTGAAAGCCGTCC 529
|
|
|
QY 161 AlaAlaSer1leAspAla1Ala1yS1yValG1yAlaAenG1yAenProLeuAenAla 180
|
|
|
Db 530 GCCGCTTCATCGACGCGGGAAGAGCCGCGTCAACAAAGCCGCTCGAGCGCC 589
|
|
|
QY 181 ValG1nG1n 183
|
|
|
Db 590 GTTCAGCAG 598
|
|
|
RESULT 3
LOCUS AR177252 638 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 1 from patent US 6312916.
ACCESSION AR177252
VERSION AR177252.1 GI:17919607
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 638)
AUTHORS Kopecky,E., Muller,R., Engh,R., Schmitz,U., Deger,A. and Brandstetter,H.
TITLE Recombinant inactive core streptavidin mutants
JOURNAL Patent: US 6312916-A 1 06-NOV-2001;
FEATURES
source 1..638
location/Qualifiers
BASE COUNT 115 a 244 c 193 g 86 t
ORIGIN /organism="unknown"

Alignment Scores:
Pred. No.: 3,24e-63 Length: 638
Score: 936.00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-589-870B-2 (1-183) x AR177252 (1-638)

QY 1 MetArgLyS1leValAla1Ala1leAlaValSerLeuThrThrValSer1leThra1a 20
|
|
|
Db 50 ATGCGCAAGATCGTCTTGACGACATCGCGCTTCCCTGACCAAGGTCTCGATTACGGCC 109
|
|
|
QY 21 SerAlaSerAlaAspProSerLyAspSerLyAsp1G1nValSerAla1aG1uAlaG1y 40
|
|
|
Db 110 AGCGCTTCGGCAGACCCCTCCAAAGACTCGAAGGCCAGGTCTCGGCCCGCGAGCGGCG 169
|
|
|
QY 41 IleThrG1yThrTyrTrpTyrAsnG1nLeuG1ySerThrPhe1leValThra1aG1yAlaAap 60
|
|
|
Db 170 ATCACCGGACACCTGTACCAACAGCTCGGCTCGACCTTCACTGTACCGCGGCGCGAGC 229
|
|
|
QY 61 G1yAla1euthrG1yThrTg1uSerAlaValG1yAenAlaG1uSerArG1yVal1leu 80
|
|
|

Db 230 GGGGCCCTGACCGGAACCTACAGAGTCGGCCGTCCGCAACCGCAGACCCCTACGTCCTG 289
|
|
|
QY 81 ThrG1yArG1yThrAspSerAlaProAlaThrAspG1ySerG1yThra1A1eug1yTTPThr 100
|
|
|
Db 290 ACCGGTCGTACGACAGCCGCCGCCAGCAGCGGCGGACCCGCTCGGTGGAGC 349
|
|
|
QY 101 ValAla1eThrLyAspAenTyrArgAenAlaHisSerAlaThrThrTPSerG1yG1nTyr 120
|
|
|
Db 350 GTGGCCCTGGAAGATACACCGCAAGCCCACTCCCGCAGCAGCTGGAGCCGCGATAC 409
|
|
|
QY 121 ValG1yG1yAlaG1uAlaArg1leAenThrG1nTTPLeuThrSerG1yThrThrG1u 140
|
|
|
Db 410 GTCCGGCGGCCCGCAGCGAGATCAACACCACTGGCTGCTACCTCCGCGACACCGAG 469
|
|
|
QY 141 AlaAenAla1eThrLySerThrLeuValG1yHisAspThrPheThrLyVal1yProSer 160
|
|
|
Db 470 GCCAAGCCTGGAAGTCACAGCTGGTGGCCACGACCTTCAACAAAGTGAAAGCCGTCC 529
|
|
|
QY 161 AlaAlaSer1leAspAla1Ala1yS1yValG1yAlaAenG1yAenProLeuAenAla 180
|
|
|
Db 530 GCCGCTTCATCGACGCGGGAAGAGCCGCGTCAACAAAGCCGCTCGAGCGCC 589
|
|
|
QY 181 ValG1nG1n 183
|
|
|
Db 590 GTTCAGCAG 598
|
|
|
RESULT 4
LOCUS AR204862 638 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 1 from patent US 6368813.
ACCESSION AR204862
VERSION AR204862.1 GI:21502296
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 638)
AUTHORS Reznik,G.O., Sano,T., Vajda,S., Smith,C. and Cantor,C.
TITLE Multifactor streptavidin
JOURNAL Patent: US 6368813-A 1 09-APR-2002;
FEATURES
source 1..638
location/Qualifiers
BASE COUNT 115 a 244 c 193 g 86 t
ORIGIN /organism="unknown"

Alignment Scores:
Pred. No.: 3,24e-63 Length: 638
Score: 936.00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-589-870B-2 (1-183) x AR204862 (1-638)

QY 1 MetArgLyS1leValAla1Ala1leAlaValSerLeuThrThrValSer1leThra1a 20
|
|
|
Db 50 ATGCGCAAGATCGTCTTGACGACATCGCGCTTCCCTGACCAAGGTCTCGATTACGGCC 109
|
|
|
QY 21 SerAlaSerAlaAspProSerLyAspSerLyAsp1G1nValSerAla1aG1uAlaG1y 40
|
|
|
Db 110 AGCGCTTCGGCAGACCCCTCCAAAGACTCGAAGGCCAGGTCTCGGCCCGCGAGCGGCG 169
|
|
|
QY 41 IleThrG1yThrTyrTrpTyrAsnG1nLeuG1ySerThrPhe1leValThra1aG1yAlaAap 60
|
|
|
Db 170 ATCACCGGACACCTGTACCAACAGCTCGGCTCGACCTTCACTGTACCGCGGCGCGAGC 229
|
|
|
QY 61 G1yAla1euthrG1yThrTg1uSerAlaValG1yAenAlaG1uSerArG1yVal1leu 80
|
|
|
Db 230 GGGGCCCTGACCGGAACCTACAGAGTCGGCCGTCCGCAACCGCAGACCCCTACGTCCTG 289
|
|
|

QY 81 ThrGlyAxyTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
DB 290 ACCGGTCGTTACGACAGCCCGCCGACCGAGCGGACCGCCCTCGGTGAGC 349
QY 101 VALAATrPlyAsnAntyArGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyr 120
DB 350 GTGGCTTGAGAAATrACTrACCGCAAGCCCACTCGGCAACGGAGCGGCAGTAC 409
QY 121 ValGlyGlyAlaGlyAlaArgIleAsnThrGlnTrpLeuThrSerGlyThrThrGlu 140
DB 410 GTCCGGGGCGCGAGCGAGATCAACACCACTGCTGACTCCGGCACCGAG 469
QY 141 AlaAsnAlaTrpLysSerThrIleuValGlyHisAspThrPheThrLysValLysProSer 160
DB 470 GCCAAGCGCTGGAAGTCCACGCTGCTGGCCACGACACCTTCACCAAGGTGAAGCGCTCC 529
QY 161 AlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsnGlyAsnProLeuAspAla 180
DB 530 GCCGCTTCATCGACCGCGGAGAAAGCGCGGTCAACACGGAACCGCTTCAGCGCC 589
QY 181 ValGlnGln 183
DB 590 GTTCAGCAG 598
RESULT 5
LOCUS AR217936 638 bp DNA linear PAT 25-SEP-2002
DEFINITION Sequence 1 from patent US 6417331.
ACCESSION AR217936
VERSION AR217936.1 GI:23318240
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 638)
AUTHORS Kopeckzki, E., Muller, R., Engh, R., Schmitt, U., Deger, A. and
TITLES Brandschetter, H.
JOURNAL Recombinant inactive core streptavidin mutants
FEATURES
source 1..638
location/Qualifiers
BASE COUNT 115 a 244 c 193 g 86 t 86 c
ORIGIN
Alignment Scores:
Pred. No.: 3,24e-63 Length: 638
Score: 936.00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-589-870B-2 (1-183) x AR217936 (1-638)
QY 1 MetArgLysIleValAlaAlaAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
DB 50 ATGGCCAGAGATCGTGGTTCAGCCCAATCGCGCTTCCCTGACCAACGAGTTCGATTACGGCC 109
QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGlyAlaGly 40
DB 110 AGCGCTTGGCAGACCCCTCCCAAGACTCGAAGGCCCGGCTCGCGCCGAGCGCCGCGC 169
QY 41 IleThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAlaAsp 60
DB 170 ATCACCGGACACTGTGTACACACAGCTCGGCTGCACTTCATGTGACCGCGGCGCCGAC 229
QY 61 GlyAlaLeuThrGlyThrTyrGluSerAlaValaGlyAsnAlaGluSerArgTyrValLeu 80
DB 230 GGCGCCCTGACCGGAGACTGAGAGTGGCGCTCGGCAACGCCGAGCGCGTACGTCCTG 289
QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100

DB 290 ACCGGTCGTTACGACAGCCCGCCGACCGAGCGGACCGCCCTCGGTGAGC 349
QY 101 VALAATrPlyAsnAntyArGAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyr 120
DB 350 GTGGCTTGAGAAATrACTrACCGCAAGCCCACTCGGCAACGGAGCGGCAGTAC 409
QY 121 ValGlyGlyAlaGlyAlaArgIleAsnThrGlnTrpLeuThrSerGlyThrThrGlu 140
DB 410 GTCCGGGGCGCGAGCGAGATCAACACCACTGCTGACTCCGGCACCGAG 469
QY 141 AlaAsnAlaTrpLysSerThrIleuValGlyHisAspThrPheThrLysValLysProSer 160
DB 470 GCCAAGCGCTGGAAGTCCACGCTGCTGGCCACGACACCTTCACCAAGGTGAAGCGCTCC 529
QY 161 AlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsnGlyAsnProLeuAspAla 180
DB 530 GCCGCTTCATCGACCGCGGAGAAAGCGCGGTCAACACGGAACCGCTTCAGCGCC 589
QY 181 ValGlnGln 183
DB 590 GTTCAGCAG 598
RESULT 6
LOCUS AX057941 638 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 1 from Patent WO0075333.
ACCESSION AX057941
VERSION AX057941.1 GI:12310569
KEYWORDS
SOURCE Streptomyces avidinii
ORGANISM Streptomyces avidinii
REFERENCE 1
AUTHORS Goshorn, S. C., Graves, S. S., Schultz, J. E., Lin, Y., Sanderson, J. A. and
TITLES Reno, J. M.
JOURNAL Streptavidin expressed gene fusions and methods of use thereof
FEATURES
source 1..638
location/Qualifiers
BASE COUNT 115 a 244 c 193 g 86 t 86 c
ORIGIN
Alignment Scores:
Pred. No.: 3,24e-63 Length: 638
Score: 936.00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-589-870B-2 (1-183) x AX057941 (1-638)
QY 1 MetArgLysIleValAlaAlaAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
DB 50 ATGGCCAGAGATCGTGGTTCAGCCCAATCGCGCTTCCCTGACCAACGAGTTCGATTACGGCC 109
QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGlyAlaGly 40
DB 110 AGCGCTTGGCAGACCCCTCCCAAGACTCGAAGGCCCGGCTCGCGCCGAGCGCCGCGC 169
QY 41 IleThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAlaAsp 60
DB 170 ATCACCGGACACTGTGTACACACAGCTCGGCTGCACTTCATGTGACCGCGGCGCCGAC 229
QY 61 GlyAlaLeuThrGlyThrTyrGluSerAlaValaGlyAsnAlaGluSerArgTyrValLeu 80
DB 230 GGCGCCCTGACCGGAGACTGAGAGTGGCGCTCGGCAACGCCGAGCGCGTACGTCCTG 289

Qy 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
Db ACCGGTCGTTACGACAGCGCCCGGCGACCGAGCGGAGCGCCCTCGTTGAGG 349
Qy 101 ValAlaTyrPlyAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyr 120
Db GTGGCTCGGAAGATTAATCAACCGCAACGCCACTCCGCGACACGCGAGCGCCAGTAC 409
Qy 121 ValGlyGlyAlaGluAlaArgGlyLeaThrGlnTyrPleuLeuThrSerGlyThrThrGlu 140
Db GTGGGGGGCGCGAGGAGATCAACCGAGTGGCTGCTGACCTCCGCGACCGAG 469
Qy 141 AlaAsnAlaTyrPlySerThrLeuValGlyHisAspThrPheThrTyrValIysProSer 160
Db GCCAAGCGCTGGAAATCCACGCTGCTGGCGCACGACCTTCCACCAAGGTGAAGCCGTCC 529
Qy 161 AlaAlaSerIleAspAlaAlaIysValIysAlaGlyValAlaAsnGlyValAsnProLeuAspAla 180
Db GCGGCTCCATCGACCGCGCGAAGAGCGCGGCTCAACAGCGCAACCGCTCGACGCC 589
Qy 181 ValGlnGln 183
Db GTTCAGCAG 598
RESULT 7
LOCUS 101349 638 bp ss-DNA linear PAT 21-MAY-1993
DEFINITION Sequence 2 from Patent US 4839293.
ACCESSION 101349
VERSION 101349.1 GI:270135
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 638)
AUTHORS Cantor, C.R., Axel, R. and Argarana, C.
TITLE DNA encoding streptavidin, streptavidin produced therefrom, fused
polypeptides which include amino acid sequences present in
streptavidin and uses thereof
JOURNAL Patent: US 4839293-A 2 13-JUN-1989;
The Trustees of Columbia University in the City of New York; New
York, NY
FEATURES
source 1..638 location/Qualifiers
BASE COUNT 114 a 244 c 194 g 86 t
ORIGIN
Alignment Scores:
Pred. No.: 5.5e-63 Length: 638
Score: 933.00 Matches: 182
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.45% Mismatches: 0
Query Match: 99.68% Indels: 0
Gaps: 0
US-09-589-870b-2 (1-183) x 101349 (1-638)

Qy 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
Db ACCGGTCGTTACGACAGCGCCCGGCGACCGAGCGGAGCGCCCTCGTTGAGG 349
Qy 101 ValAlaTyrPlyAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyr 120
Db GTGGCTCGGAAGATTAATCAACCGCAACGCCACTCCGCGACACGCGAGCGCCAGTAC 409
Qy 121 ValGlyGlyAlaGluAlaArgGlyLeaThrGlnTyrPleuLeuThrSerGlyThrThrGlu 140
Db GTGGGGGGCGCGAGGAGATCAACCGAGTGGCTGCTGACCTCCGCGACCGAG 469
Qy 141 AlaAsnAlaTyrPlySerThrLeuValGlyHisAspThrPheThrTyrValIysProSer 160
Db GCCAAGCGCTGGAAATCCACGCTGCTGGCGCACGACCTTCCACCAAGGTGAAGCCGTCC 529
Qy 161 AlaAlaSerIleAspAlaAlaIysValIysAlaGlyValAlaAsnGlyValAsnProLeuAspAla 180
Db GCGGCTCCATCGACCGCGCGAAGAGCGCGGCTCAACAGCGCAACCGCTCGACGCC 589
Qy 181 ValGlnGln 183
Db GTTCAGCAG 598
RESULT 8
LOCUS 578777 625 bp DNA linear BCT 30-OCT-1995
DEFINITION streptavidin v1 [Streptomyces violaceus, Tu 2460, Genomic, 625 nt].
ACCESSION 578777
VERSION 578777.1 GI:1042193
KEYWORDS
SOURCE Streptomyces violaceus
ORGANISM Streptomyces violaceus
REFERENCE 1 (bases 1 to 625)
AUTHORS Bayer E.A., Kulik, T., Adar, R. and Wlchek, M.
TITLE Close similarity among streptavidin-like, biotin-binding proteins
from Streptomyces
JOURNAL Biochim. Biophys. Acta 1263 (1), 60-66 (1995)
MEDLINE 7632734
PUBMED 95359204
REMARK Genbank staff at the National Library of Medicine created this
entry [NCBI gi2460.169185] from the original journal article.
This sequence comes from Fig. 4.
FEATURES
source 1..625 location/Qualifiers
BASE COUNT 115 a 240 c 186 g 84 t
ORIGIN
Alignment Scores:
Pred. No.: 6.41e-63 Length: 625
Score: 932.00 Matches: 182
Percent Similarity: 99.45% Conservative: 0
Best Local Similarity: 99.45% Mismatches: 1
Query Match: 99.57% Indels: 0
/organism="Streptomyces violaceus"
/mol_type="genomic DNA"
/db_xref="taxon:1936"
50..601
/gene="streptavidin v1, Sa v1"
50..601
/gene="streptavidin v1, Sa v1"
/note="This sequence comes from Fig. 5; Sa v1"
/codon_start=1
/product="streptavidin v1"
/protein_id="AAB35015.1"
/db_xref="GI:1042194"
/translation="MKRIIVAAIVSLTTSITASADPSKDSKAOVSAAEAGITGT
WKNINRNHSAITTSAGTVGTEARINIMULTISGTETANMKSTLVGHDFTKVPS
AASIDPAKKAQVNNGNPDAVQO"
CDS
gene
CDS

DB: 1 Gaps: 0
US-09-589-870b-2 (1-183) x S78777 (1-625)
QY 1 MetArgLysIleValAlaIleAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
Db 50 ATGCGCAAGATCGTGTGAGCCATCGCGTTTCCCTGACCGGCTCGATTACCGCC 109
QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaIleAlaGlnVal 40
Db 110 AGGCGCTTCGCGCAACCCCTCCAGACTCGAAGGCCAGGTCTCGGCCCGCGAGCCGGC 169
QY 41 IleThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValThrAlaGlnValAsp 60
Db 170 ATCACCAGCGACCTGCTGACCAACAGCTCGGCTCGACCTTCATCGTACCGCGGCGCCGAC 229
QY 61 GlnValLeuThrGlyThrTyrGlnSerAlaValAlaGlnValAsnAlaGlnSerArgTyrValLeu 80
Db 230 GGGCGCCCTGACCGGAACCTACGAGCGCGCTCGGCAAGCGCCGAGAGCGGCTACGCTCG 289
QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
Db 290 ACCGCTGTTACGACAGCGCGCCCGGACCGACCGAGCGGACCGCCCTCGGCTGAGACG 349
QY 101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyr 120
Db 350 GTGGCTGGAAGATTAATACCGCAACGCGCACTCCGACGACGAGCGGAGCGGCGAGTAC 409
QY 121 ValGlyGlyAlaGlnValArgIleAsnThrGlnThrLeuLeuThrSerGlyThrThrGlu 140
Db 410 GTCCGGGACCGAGCGGAGATCAACACCGAGTGTCTGACCTCGCGGACCGACCGAG 469
QY 141 AlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValLysProSer 160
Db 470 GCCAAGCGCTGGAAGTCCAGCGTGTGCGGACGACCACTTCACCAAGTGAAGCCGTCC 529
QY 161 AlaAlaSerIleAspAlaAlaLysLysValAlaGlnValAsnAsnGlyAsnProLeuAspAla 180
Db 530 GCGCGCTTCATCGACCGCGGAGAAAGCGCGGCTGTCMACGAGCAACCGCTCGACGCC 589
QY 181 ValGlnGln 183
Db 590 GTTCAGCAG 598
RESULT 9
LOCUS S78782 625 bp DNA linear BCT 30-OCT-1995
DEFINITION streptavidin v2 [Streptomyces violaceus, Tu 2605, Genomic, 625 nt].
ACCESSION S78782
VERSION S78782.1 GI:1042195
KEYWORDS
SOURCE
ORGANISM Streptomyces violaceus
Streptomyces violaceus
Bacteria; Actinobacteriae; Actinomycetales;
Streptomycetaceae; Streptomyces.
REFERENCE
AUTHORS 1 (bases 1 to 625)
TITLE Beyer, E.A., Kolik, T., Adar, R. and Wilchek, M.
JOURNAL Close similarity among streptavidin-like, biotin-binding proteins
MEDLINE from Streptomyces
PUBMED Biochim. Biophys. Acta 1263 (1), 60-66 (1995)
REMARK 7637734
Genbank scaff at the National Library of Medicine created this
entry [NCBI gi169186] from the original journal article.
This sequence comes from Fig. 4.
FEATURES
source
1..625
Location/Qualifiers
/organism="Streptomyces violaceus"
/mol_type="genomic DNA"
/db_xref="taxon:1936"
50..601
/gene="streptavidin v2, Sa v2"
50..601
CDS

/gene="streptavidin v2, Sa v2"
/note="This sequence comes from Fig. 5; Sa v2"
/codon_start=1
/product="streptavidin v2"
/protein_id="BAB35016.1"
/db_xref="GI:1042196"
translation="MRKIVVAIAVSLTGTITASASADPSKDAQAQAAVAEAGTGT
WYNOLGTFEIVTANADSLTGTVSAVNAASRYVLGRVDSAPATDGGTALCMIVA
WKNVRNHSATITWSGQIVASSEARINTQMLITGTTAANMKSTLVGHDTFTKVPK
AASIDPAKKAGVNGNPUDAVQ"

BASE COUNT 117 a 244 c 184 g 80 t

ORIGIN

Alignment Scores:
Pred. No.: 3,04e-60 Length: 625
Score: 897.00 Matches: 174
Percent Similarity: 96.72% Conservative: 3
Best Local Similarity: 95.08% Mismatches: 6
Query Match: 95.83% Indels: 0
DB: 1 Gaps: 0

US-09-589-870b-2 (1-183) x S78782 (1-625)
QY 1 MetArgLysIleValAlaIleAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
Db 50 ATGCGCAAGATCGTGTGAGCCATCGCGTTTCCCTGACCGGCTCGATTACCGCC 109
QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaIleAlaGlnVal 40
Db 110 AGGCGCTTCGCGCAACCCCTCCAGACTCGAAGGCCAGGTCTCGGCCCGCGAGCCGGC 169
QY 41 IleThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValThrAlaGlnValAsp 60
Db 170 ATCACCAGCGACCTGCTGACCAACAGCTCGGCTCGACCTTCATCGTACCGCGGCGCCGAC 229
QY 61 GlnValLeuThrGlyThrTyrGlnSerAlaValAlaGlnValAsnAlaGlnSerArgTyrValLeu 80
Db 230 GGGCGCCCTGACCGGAACCTACGAGCGCGCTCGGCAAGCGCCGAGAGCGGCTACGCTCG 289
QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
Db 290 ACCGCTGTTACGACAGCGCGCCCGGACCGACCGAGCGGACCGCCCTCGGCTGAGACG 349
QY 101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyr 120
Db 350 GTGGCGTGAAGAAACCACTACCGGACCGCCACCTCCGACGACGAGCGGCACTAC 409
QY 121 ValGlyGlyAlaGlnValArgIleAsnThrGlnThrLeuLeuThrSerGlyThrThrGlu 140
Db 410 GTCCGGGACCGAGCGGAGATCAACACCGAGTGTCTGACCTCGCGGACCGACCGAG 469
QY 141 AlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValLysProSer 160
Db 470 GCCAAGCGCTGGAAGTCCAGCGTGTGCGGACGACCACTTCACCAAGTGAAGCCGTCC 529
QY 161 AlaAlaSerIleAspAlaAlaLysLysValAlaGlnValAsnAsnGlyAsnProLeuAspAla 180
Db 530 GCGCGCTTCATCGACCGCGGAGAAAGCGCGGCTGTCMAAGCAACCGCTCGACGCC 589
QY 181 ValGlnGln 183
Db 590 GTTCAGCAG 598
RESULT 10
LOCUS AX057945 1239 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 5 from Patent WO0075333.
ACCESSION AX057945
VERSION AX057945.1 GI:12310571
KEYWORDS
SOURCE Streptomyces avidinii
ORGANISM Streptomyces avidinii
Bacteria; Actinobacteriae; Actinomycetales;

REFERENCE		Streptomyces; Streptomycetaceae; Streptomycetes.
AUTHORS	1	Goshorn, S.C., Gravel, S.S., Schultz, J.E., Lin, Y., Sanderson, J.A. and Remo, J.M.
TITLE		Streptavidin expressed gene fusions and methods of use thereof
JOURNAL		Patent: WO 0075333-A 5 14-DEC-2000;
FEATURES		NEORX CORPORATION (US)
SOURCE		Location/Qualifiers 1..1239 /organism="Streptomyces avidinii" /mol_type="genomic DNA" /db_xref="taxon:1895"
BASE COUNT	270 a 392 c 356 g 221 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	9,67e-57	Length: 1239
Score:	855.50	Matches: 165
Percent Similarity:	98.83%	Conservative: 4
Best Local Similarity:	96.49%	Mismatches: 1
Query Match:	91.40%	Indels:
DB:	6	Gaps: 1
US-09-589-870B-2 (1-183) x AX057945 (1-1239)		
OY	14 ThrThrValSerIleThrAla---SerAlaSerAlaAppProSerLyAspSerLysAla	32
Db	724 ACCACGGCTCACCGTGAAGCTGCTGTGCTGTTGGCAGACCCTCCCAAGACTCGAAGGCC	783
OY	33 GlnValSerAlaAlaGluAlaGlyLyleThrGlyThrTrpTyranGlnLeuGlySerThr	52
Db	784 CAGGCTCTGGCCGCCGAGGCGCGGCATCCAGGACCTGGTACAACAGCTGGCTTCACC	843
OY	53 PheIleValIleThrAlaGlyAlaAspDylValLeuThrGlyThrTyrglusSerAlaValGly	72
Db	844 TTCATCGTGACCGCGCGGCCGCGACGCCCTTGACCGGAACCTTAGAGTGGCCCTGGCC	903
OY	73 AsnAlaGluSerArgTyValleuThrGlyArgTyraPserAlaProAlaThrAspGly	92
Db	904 AACCGCGAGACCGCTGACGCTCTGCTGACCGCGCTTAACGACGCGCCGCCACCGAGCGC	963
OY	93 SerGlyThrAlaLeuGlyTyrThrValAlaIleTrpLysAsnAsnTyraGlnAlaHisSer	112
Db	964 AGCGGACCGCCCTCGGTGGACCGGTGGCCCTGGAAGAATACTACCCGACCCACTCC	1023
OY	113 AlaThrThrTPSerGlyGlnTyraValGlyGluAlaGluAlaArgIleAsnThrGlnTrp	132
Db	1024 GCGACCACTGAGAGCGGCCAGTACCTCGCGCGCGCGGAGGAGGAGATCAACACCCAGTGG	1083
OY	133 LeuLeuThrSerGlyThrThGluAlaAsnAlaIleTrpLysSerThrLeuValGlyHisAsp	152
Db	1084 CTGCTGACCTCCGGCACCAACGAGGCCAACGCCCTGGAAGTCCAAGCTGCTGGCCACGAC	1143
OY	153 ThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyVal	172
Db	1144 ACCTTACCAAGGATGAAGCCCTCGCGCCCTTCATCGACGCGGCGAAGAAGGCCGCGGTC	1203
OY	173 AsnAsnGlyAsnProLeuAspAlaValGlnGln	183
Db	1204 AACAAAGGCAACCCGCTGACGCGCTTACGAG	1236
RESULT 11		
AX057943		
LOCUS	AX057943	1612 bp DNA linear PAT 17-JAN-2001
DEFINITION	Sequence 3 from Patent W0075333.	
ACCESSION	AX057943	
VERSION	AX057943.1	GI:12310570
KEYWORDS		
SOURCE		Streptomyces avidinii
ORGANISM		Streptomyces avidinii
REFERENCE		Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

AUTHORS	Geshorn, S. C., Graves, S. S., Schultz, J. E., Lin, Y., Sanderson, J. A. and Reno, D. M.
TITLE	Streptavidin expressed gene fusions and methods of use thereof
JOURNAL	Patent: WO 0075333-A3 14-DEC-2000; NEORX CORPORATION (US)
FEATURES	location/Qualifiers
SOURCE	1..1612 /organism="Streptomyces avidinii" /mol_type="genomic DNA" /db_xref="taxon:1895"
BASE COUNT	353 a 493 c 459 g 307 t
ORIGIN	
Alignment Scores:	
Pred. No.:	1,54e-56 Length: 1612
Score:	854.50 Matches: 166
Percent Similarity:	97.67% Conservative: 2
Best Local Similarity:	96.51% Mismatches: 3
Query Match:	91.29% Indels: 1
DB:	Gaps: 1
US-09-589-870B-2 (1-183) x AX057943 (1-1612)	
QY	12 SerleuThrThrValSerIleThrAlaSerAlaSerAlaAspProSerIleAspSerIle 31
DB	1091 ACCTTAGTCACCGGAGAC---TCGGCTCGGTTCCGCAACCCCTCCAGACTCCAG 1147
QY	32 AlaGlnValSerAlaAlaGluAlaGlyIleThrGlyThrIlePheGlnIleGlySer 51
DB	1148 GCCACGCTCTCGCGCCGCGGCGGCGGATACCGGACCTCGTACCAACGACTCGCTCG 1207
QY	52 ThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrIleGluSerAlaVal 71
DB	1208 ACCTTCATCGGACCGCGGGCGGCGGACGGCGCCCTGACCGGAACCTACGATCGCGCTC 1267
QY	72 GlyAlaAlaGluSerAlaGlyIleValLeuThrGlyAlaGlyIleAspSerAlaIleThrAsp 91
DB	1268 GGCACACCCGAGACCCCTCAGTCTGACCGGTCGTACGACAGCGCCCGGCGCACCGAC 1327
QY	92 GlySerGlyThrAlaLeuGlyIleThrThrValAlaIleThrPheAsnAlaIlePheArgAsnAlaHis 111
DB	1328 GGCAGCGGCACCGCCCTCGGTTGACGCGGTGGCGCTCGGAAGAAATACCTACCGACAGCCAC 1387
QY	112 SerAlaThrThrIlePheSerGlyGlnIleValGlyIleAlaGluAlaArgIleAsnThrGln 131
DB	1388 TCCGCGACCAAGTGGACGGCGCAAGTACGTGGGGGGCGCGGCGGAGATCAACACCCAG 1447
QY	132 TrpLeuLeuThrSerGlyIleThrGluAlaAsnAlaTrpIleSerThrLeuValGlyHis 151
DB	1448 TGGCTGCTGACCTCCGCGACACCGAGGCCACGCCCTGGAAGTCCACGCTGCTCGCGCAC 1507
QY	152 AspThrPheThrIleValIleValProSerAlaAlaSerIleAspAlaAlaValIleValAlaGly 171
DB	1508 GACACCTTACCAAGGAGGAGCGGTCCGCCCTTCATCCAGCGGGGAGGAGAGCGCGGC 1567
QY	172 ValAsnAsnGlyAsnProLeuAspAlaValAlaGlnGln 183
DB	1568 GTCAACAACGGCAACCCGCTCGACGCCGTTCAACAG 1603
RESULT 12	
AX057947	
LOCUS	AX057947 1280 bp DNA linear PAT 17-JAN-2001
DEFINITION	Sequence 7 from Patent WO0075333.
ACCESSION	AX057947
VERSION	AX057947.1 GI:12310572
KEYWORDS	
SOURCE	Streptomyces avidinii
ORGANISM	Streptomyces avidinii
REFERENCE	Bacteria, Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
AUTHORS	1 Goehorn, S. C., Graves, S. S., Schultz, J. E., Lin, Y., Sanderson, J. A. and Reno, J. M.

TITLE Streptavidin expressed gene fusions and methods of use thereof
JOURNAL Patent: WO 0075333-A 7 14-DEC-2000;
NEORX CORPORATION (US)
FEATURES Location/Qualifiers
source 1..1280
/organism="Streptomyces avidinii"
/mol_type="genomic DNA"
/db_xref="taxon:1895"
BASE COUNT 267 a 397 c 388 g 228 t
ORIGIN

Alignment Scores:
Pred. No.: 4,48e-56 Length: 1280
Score: 847.00 Matches: 162
Percent Similarity: 98.79% Conservative: 1
Best Local Similarity: 98.18 Mismatches: 2
Query Match: 90.49% Indels: 0
DB: 6 Gaps: 0

US-09-589-870b-2 (1-183) x AX057947 (1-1280)

QY 19 ThrAlaSerAlaSerAlaSPProSerLyAspSerLyAlaGlnValSerAlaAlaGln 38
Db 777 TCTGGCTCTGCTTGGCGACACCCCTCCAGAGACTCGAAGGCCAGGCTCTGGCCCGCGAG 836
QY 39 AlaGlyIleThrGlyThrTPYrAsnGlnLeuGlySerThrPheIleValThrAlaGly 58
Db 837 GCGGCGATCACCGGCACCTGTTACACACGCTCGGCTCGACCTTCCTCGACCGCGGCG 896
QY 59 AlaAspGlyAlaLeuThrGlyThrTPYrGlnSerAlaValGlyAsnAlaGlnSerArgTyr 78
Db 897 GCCGAGCGGCGCTGACCGGAACCTACGAGTCGGCGCGGCAACGCCGAGAGCGCTAC 956
QY 79 ValLeuThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGly 98
Db 957 GTCTGACCGGCTGTTACACACGCGCCGCGCACCGAGCGAGCGGCGACCGCCCTCGGT 1016
QY 99 ThrPheValAlaThrLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTPSerGly 118
Db 1017 TGGACGGTGGCTCGGAAGATTACTACCGCAAGCCCACTCGCGCACCGTGGAGCGGCG 1076
QY 119 GlnTyrValGlyGlyAlaGlnAlaArgIleAsnThrGlnThrLeuLeuThrSerGlyThr 138
Db 1077 CAGTACGTGGCGGCGCCGAGCGAGATCAACACCACTGCTGTGACCTCCGCGCAC 1136
QY 139 ThrGlnAlaAsnAlaThrLysSerThrLeuValGlyHisAspThrPheThrLysValLys 158
Db 1137 ACCGAGGCCAACGCTCGAAGCTCCAGCTGCTGCGCCACGACACCTTCACCAAGGTGAAG 1196
QY 159 ProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsnGlyAsnProLeu 178
Db 1197 CCGTCGCGCGCTCCATCGACGCGGAGAAAGCGCGCTCAACCAACGCAACCGGCTC 1256
QY 179 AspAlaValGlnGln 183
Db 1257 GACGCCGTTACAGAG 1271

RESULT 13
A39565 1356 bp DNA linear PAT 05-MAR-1997
LOCUS Sequence 4 from Patent WO9415644.
DEFINITION A39565
ACCESSION A39565.1 GI:2295847
VERSION A39565.1 GI:2295847
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1356)
AUTHORS Epenetos,A.A., Spooner,R.A. and Deonaraiah,M.
TITLE COMPOUNDS FOR TARGETING
JOURNAL Patent: WO 9415644-A 4 21-JUL-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication GB 2289679 951129.

FEATURES Location/Qualifiers
source 1..1356
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
40..1347
/note="unnamed protein product"
/codon_start=1
/protein_id="CA02449.1"
/db_xref="GI:2295848"
translation="MKYLLPTAAGLLILAAOPAAVOYLQOPGAELVPGASVYLSCKASGTYPTSYMMHWKORPRGLENIGIDPNSGCTKNEKFKSKATLTVPKSTAYMQLSLTSEDAVYICARYDYGSYFDWQGVTVSVSGSGSGGGSGGSGQAVVTOESALTPSPGETVTLTCRSSTGAVTTSNANWQEPDLFTGLIGTNNRAPGVARFSGSLTIGDKAALITGAOTEDALFYCALMYSNHWYFGGTXLTVLGLAPAAPADPSKSKQAVAAEAGITGWNOGSEFFIVTAGADGALGTYESAVGNASRYVLTGRYDAPATDGGSTALGWTVMKKNNRYRNATWVSGVYGAERINQWLLTSGTTEANMKSTLVGHDITFKVPSAASIDAAKKAGVNNGNPLDANVQ"

BASE COUNT 332 a 357 c 361 g 306 t
ORIGIN

Alignment Scores:
Pred. No.: 5,69e-56 Length: 1356
Score: 846.00 Matches: 162
Percent Similarity: 99.39% Conservative: 1
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 90.38% Indels: 0
DB: 6 Gaps: 0

US-09-589-870b-2 (1-183) x A39565 (1-1356)

QY 20 AlaSerAlaSerAlaSPProSerLyAspSerLyAlaGlnValSerAlaAlaGln 39
Db 853 GCTGCGGACCTCGACGAGACCCCTCCAGAGACTCAAGCTCAGGTTCTGACGCCAAGCT 912
QY 40 GlyIleThrGlyThrTPYrAsnGlnLeuGlySerThrPheIleValThrAlaGlyVal 59
Db 913 GGTATCACTGGACACTGGATTAACCACTGGGGCTGCACTTCATTGTGACCGCTGGTGG 972
QY 60 AspGlyAlaLeuThrGlyThrTPYrGlnSerAlaValGlyAsnAlaGlnSerArgTyrVal 79
Db 973 GACGAGCTCTGACCTCGGACCTAGCAATCTCGGTGTAACGCAAGTCCCGCTACGTA 1032
QY 80 LeuThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyr 99
Db 1033 CTGACTGGCCGCTTATGACTCTGACCTCCACCAAGTGGCTCGTAACCGCTCTGGGCTGG 1092
QY 100 ThrValAlaThrLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTPSerGlyGln 119
Db 1093 ACTGTGGCTTGGAAAAACAACCTATCGTAATCGGCACGCGCACCTACGTGTGGCCAA 1152
QY 120 TyrValGlyGlyAlaGlnAlaArgIleAsnThrGlnThrLeuLeuThrSerGlyThrThr 139
Db 1153 TACGTGGCGGCTCGAAGCTCTGATCAACCTACAGTGGCTGTTAACATCCGGGCACTAAC 1212
QY 140 GlnAlaAsnAlaThrLysSerThrLeuValGlyHisAspThrPheThrLysValLysPro 159
Db 1213 GAAGCAATGATGAGAAATTCACACTAGTACGTATGACACCTTTACCAAGATTAAAGCT 1272
QY 160 SerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsnGlyAsnProLeuAsp 179
Db 1273 TCTGCTGTACATGATGATGTGCCAAGAACGAGCGTAAACACAGTAAACCTCTTAGAC 1332
QY 180 AlaValGlnGln 183
Db 1333 GCTGTTCAGCAA 1344

RESULT 14
A39152 1356 bp DNA linear PAT 22-JAN-2000
LOCUS A39152
DEFINITION Sequence 4 from Patent EP0815872.
ACCESSION A39152
VERSION A39152.1 GI:6741540

KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1356)
AUTHORS Epenetos, A.A. and Deonaraín, M.
TITLE Compounds for targeting
JOURNAL Patent: EP 0815872-A 4 07-JAN-1998;
IMP CANCER RES TECH (GB)
FEATURES
source Location/Qualifiers
1..1356
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
40..1347
/note="unnamed protein product"
/codon_start=1
/protein_id="CAB69540.1"
/db_xref="GI:6741541"
/translation="MKYLLPTAAAGLLLLAOPAPMAQVLOQPGAEIVKSGAVKLSK
KASGYFTSYMMHMKVQKRGLEWIGRIIDPSGCTKVKRKSATLTVDPSSSTAY
MGLSLTSDSAVYVCARDYDYGSSIFDYWGCTTVTVSSGGSGGGGGGGGGGGGG
VTQESALITTSPEETVLTICRSTGAVTTSNVANWQEKPDHFTGLIGSTNNRPGVP
ARFSGSLIDKALITTTGAQTEDEALVPCALWYSNMVGGCTKLTVLGLAEPAPAPA
DSKDSKAQVSAAEAGITGTWYQLGSTIVTAGAGALGTYESAVGNAESRYVLTG
RYDSAPATDGGSGTALGWTVAAMKNYNNANSAITWGSQVYGGAEARINTQWLLTSGTTE
ANAKSTLVGHDTFTKVRSAASIDAKKAGVNNNGPLDAVQ"

BASE COUNT 332 a 357 c 361 g 306 t
ORIGIN

Alignment Scores:
Pred. No.: 5,698-56 Length: 1356
Score: 846.00 Matches: 162
Percent Similarity: 99.39% Conservative: 1
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 90.38% Indels: 0
DB: 6 Gaps: 0

US-09-589-870b-2 (1-183) x A93152 (1-1356)

QY 20 AAlaSerAlaSerAlaApProSerLyAspSerLyAlaGlnValSerAlaAlaGlnAla 39
DB 853 GCTGCCGACCTGCACACCCGTCACAGACTCCAAAGCTCAGGTTCTGCAGCCGAACT 912

QY 40 GYIIEthGlyThrTrpTyraGlnGlnGlySerThrPheIleValThrAlaGlyAla 59
DB 913 GGATCTACCTGGCACCCTGATTAACCACTGGGGCTGACCTTCAATGTGACCGCTGTGG 972

QY 60 AaPGIyAlaLeuThrGlyThrTyraGlnGlnGlySerThrPheIleValThrAlaGlyAla 79
DB 973 GACGGAGCTCTGACCTGACCTTACGAATCTGGCTGGTGAACCGCAATCCCGCTACGTA 1032

QY 80 LeuThrGlyArgTyraPserAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyr 99
DB 1033 CTGACTGGCCGTTATGACTCTGCACCTGCACGATGGCTGTGCTGCTGCGGCTGG 1092

QY 100 ThrValAlaATrPlyAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGln 119
DB 1093 ACTGGGCTTGGAAAAACAATCTGTAATGGCCACAGCGCACTACGCTGTGGCCAA 1152

QY 120 TyrValAlaGlyAlaGlnAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThr 139
DB 1153 TACGTTGGCGGCTGAGGCTGCTATCAACACTCACTGAGCTGTAAACATCCGGCACTACC 1212

QY 140 GluAlaAsnAlaTrpLySerThrLeuValGlyHisAspThrPheThrLyValIlyPro 159
DB 1213 GAAGCAATGATGAAATCGACACTAGTCACTGACACCTTTACCAAGTTAAAGCTT 1272

QY 160 SerAlaAlaSerIleAspAlaAlaIlyValIlyValAsnAsnGlyAsnProLeuAsp 179
DB 1273 TCTGCTGTAGCATTTGATGCTGCCAAGAAAGAGGCTAAACACCGTAACCTCTAGAC 1332

QY 180 AlaValGlnGln 183

DB 1333 GCTGTTCAAGCA 1344

RESULT 15
AR082490 1356 bp DNA linear PAT 31-AUG-2000
LOCUS AR082490
DEFINITION Sequence 4 from patent US 5973116.
ACCESSION AR082490
VERSION AR082490.1 GI:1009216
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1356)
AUTHORS Epenetos, A.Antonion., Spooner, R.Anthony. and Deonaraín, M.
TITLE Compounds for targeting
JOURNAL Patent: US 5973116-A 4 26-OCT-1999;
FEATURES
source Location/Qualifiers
1..1356
/organism="unknown"

BASE COUNT 332 a 357 c 361 g 306 t
ORIGIN

Alignment Scores:
Pred. No.: 5,698-56 Length: 1356
Score: 846.00 Matches: 162
Percent Similarity: 99.39% Conservative: 1
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 90.38% Indels: 0
DB: 6 Gaps: 0

US-09-589-870b-2 (1-183) x AR082490 (1-1356)

QY 20 AAlaSerAlaSerAlaApProSerLyAspSerLyAlaGlnValSerAlaAlaGlnAla 39
DB 853 GCTGCCGACCTGCACACCCGTCACAGACTCCAAAGCTCAGGTTCTGCAGCCGAACT 912

QY 40 GYIIEthGlyThrTrpTyraGlnGlnGlySerThrPheIleValThrAlaGlyAla 59
DB 913 GGATCTACCTGGCACCCTGATTAACCACTGGGGCTGACCTTCAATGTGACCGCTGTGG 972

QY 60 AaPGIyAlaLeuThrGlyThrTyraGlnGlnGlySerThrPheIleValThrAlaGlyAla 79
DB 973 GACGGAGCTCTGACCTGACCTTACGAATCTGGCTGGTGAACCGCAATCCCGCTACGTA 1032

QY 80 LeuThrGlyArgTyraPserAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyr 99
DB 1033 CTGACTGGCCGTTATGACTCTGCACCTGCACGATGGCTGTGCTGCTGCGGCTGG 1092

QY 100 ThrValAlaATrPlyAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGln 119
DB 1093 ACTGGGCTTGGAAAAACAATCTGTAATGGCCACAGCGCACTACGCTGTGGCCAA 1152

QY 120 TyrValAlaGlyAlaGlnAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThr 139
DB 1153 TACGTTGGCGGCTGAGGCTGCTATCAACACTCACTGAGCTGTAAACATCCGGCACTACC 1212

QY 140 GluAlaAsnAlaTrpLySerThrLeuValGlyHisAspThrPheThrLyValIlyPro 159
DB 1213 GAAGCAATGATGAAATCGACACTAGTCACTGACACCTTTACCAAGTTAAAGCTT 1272

QY 160 SerAlaAlaSerIleAspAlaAlaIlyValIlyValAsnAsnGlyAsnProLeuAsp 179
DB 1333 GCTGTTCAAGCA 1344

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: October 28, 2003, 01:30:48 ; Search time 248 Seconds
(without alignments)
1978.910 Million cell updates/sec

Title: US-09-589-870B-2

Perfect score: 936
Sequence: 1 MKRIVVAIAVSLTVSITA.....IDAKKAGVGNNGPLDAVQ 183

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delpop 6.0 , Delpext 7.0	

Searched: 1792395 seqs, 1340900451 residues

Total number of hits satisfying chosen parameters: 3584790

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+.p2n.model -DEV=xlh
-Q/cg2.1/USPTO.epool/US09589870/rnuc_27102003_104450_16725/app.query.fasta_1.327
-DB=Published.Applications.NA -QFMT=fastcap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS-human0.cgi -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEARSIZE=500 -MILEN=0
-MAXLEN=200000000 -USER=US09589870.QCGN_1_1221@rnuc_27102003_104450_16725
-NCPUE=0 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FAPOP=6 -FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELXT=7

Database : Published Applications_NA.*

1:	/cg2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2:	/cg2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3:	/cg2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4:	/cg2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5:	/cg2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6:	/cg2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7:	/cg2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8:	/cg2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9:	/cg2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10:	/cg2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11:	/cg2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
12:	/cg2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13:	/cg2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
14:	/cg2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
15:	/cg2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
16:	/cg2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
17:	/cg2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	936	100.0	638	12	US-10-244-821-1 Sequence 1, Appli

2	936	100.0	638	14	US-10-013-173-1	Sequence 1, Appli
3	936	100.0	638	14	US-10-150-762-1	Sequence 1, Appli
4	855.5	91.4	1239	12	US-10-244-821-5	Sequence 5, Appli
5	855.5	91.4	1239	14	US-10-013-173-5	Sequence 5, Appli
6	855.5	91.4	1239	14	US-10-150-762-5	Sequence 5, Appli
7	854.5	91.3	1614	12	US-10-244-821-3	Sequence 3, Appli
8	854.5	91.3	1614	14	US-10-013-173-3	Sequence 3, Appli
9	854.5	91.3	1614	14	US-10-150-762-3	Sequence 3, Appli
10	847	90.5	1280	12	US/10/244	Sequence 7, Appli
11	847	90.5	1280	14	US/10/013	Sequence 7, Appli
12	847	90.5	1280	14	US/10/150	Sequence 7, Appli
13	847	90.5	1271	12	US-10-244-821-87	Sequence 87, Appli
14	847	90.5	1467	12	US-10-244-821-48	Sequence 48, Appli
15	847	90.5	1467	14	US-10-013-173-48	Sequence 48, Appli
16	847	90.5	1467	14	US-10-150-762-48	Sequence 48, Appli
17	834	89.1	498	10	US-09-111-447-7	Sequence 7, Appli
18	834	89.1	1266	10	US-09-938-2708-2	Sequence 2, Appli
19	687.5	73.5	1173	12	US-10-075-947A-4	Sequence 4, Appli
20	682	72.9	1176	12	US-10-075-947A-3	Sequence 3, Appli
21	110.5	11.8	3381	11	US-09-952-267-6	Sequence 6, Appli
22	105.5	11.3	3018	12	US-10-193-764-40	Sequence 40, Appli
23	105.5	11.3	3036	12	US-10-193-764-38	Sequence 38, Appli
24	102.5	11.0	3331	9	US-09-864-761-19481	Sequence 19481, A
25	101.5	10.8	3300	9	US-09-379-931-6	Sequence 6, Appli
26	101.5	10.8	3300	12	US-10-223-597-6	Sequence 6, Appli
27	99.5	10.6	3108	12	US-10-193-764-58	Sequence 68, Appli
28	99.5	10.6	4837	12	US-10-193-764-56	Sequence 66, Appli
29	99.5	10.6	4937	13	US-10-092-880-3	Sequence 3, Appli
30	99.5	10.6	7407	12	US-10-246-310-3	Sequence 3, Appli
31	99.5	10.6	9323	13	US-10-092-880-6	Sequence 6, Appli
32	98	10.5	3349	11	US-09-952-267-2	Sequence 2, Appli
33	97.5	10.4	489	9	US-09-864-761-27528	Sequence 27528, A
34	96.5	10.3	1089	14	US-10-156-761-3313	Sequence 3313, Ap
35	96.5	10.3	2247	11	US-09-272-975-3	Sequence 3, Appli
36	96.5	10.3	7507	11	US-09-272-975-1	Sequence 1, Appli
37	96.5	10.3	902608	14	US-10-156-761-1	Sequence 1, Appli
38	95	10.1	902608	14	US-10-156-761-1	Sequence 1, Appli
39	94	10.0	801	10	US-09-974-300-6560	Sequence 6560, Ap
40	93.5	10.0	896	12	US-10-101-510-732	Sequence 732, Ap
41	93.5	10.0	1329	14	US-10-156-761-6352	Sequence 6352, Ap
42	92.5	9.9	825	14	US-10-156-761-7318	Sequence 7318, Ap
43	92.5	9.9	3711	11	US-09-769-787-321	Sequence 321, Ap
44	92.5	9.9	25165	14	US-10-114-170-39	Sequence 39, Appli
45	92	9.8	461	9	US-09-864-761-10890	Sequence 10890, A

ALIGNMENTS

RESULT 1
US-10-244-821-1
Sequence 1, Application US/10244821
Publication No. US2003014323A1
GENERAL INFORMATION:
APPLICANT: Goshorn, Stephen Charles
APPLICANT: Graves, Scott Stoll
APPLICANT: Schultz, Joanne Elaine
APPLICANT: Lin, Yukang
APPLICANT: Sanderson, James Allen
APPLICANT: Reno, John M.
APPLICANT: Dearstine, Erica A.
TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
TITLE OF INVENTION: METHODS OF USE THEREOF
FILE REFERENCE: 690022.547C3
CURRENT APPLICATION NUMBER: US/10/244,821
CURRENT FILING DATE: 2002-09-16
NUMBER OF SEQ ID NOS: 92
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 638
TYPE: DNA
ORGANISM: Streptomyces avidinii
US-10-244-821-1

Alignment Scores:

Pred. No.: 1,92e-101 Length: 638
Score: 936.00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-09-589-870B-2 (1-183) x US-10-244-821-1 (1-638)

```
QY 1 MetArgLysIleValAlaAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
    |||
DB 50 ATGCGAAGATCGTCGTCGAGCCATCGCGTTCCCTGACCAACGCGTCTCGATTACGGCC 109
QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGlnAgi 40
    |||
DB 110 AGCGCTTCGGCAGACCCCTCCAGAGACTCGAAGGCCAGGTCTCGGCCCGCCAGGCCGC 169
QY 41 IleThrGlyThrTyrTyrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAla 60
    |||
DB 170 ATCACCAGGACCTGTTACACACAGCTCGGCTCGACTTCATCGTACCGCGGCGCGAC 229
QY 61 GlyAlaLeuThrGlyThrTyrGlySerAlaValGlyAsnAlaGlySerArgTyrValLeu 80
    |||
DB 230 GCGGCCCTGACCGAACCCTACGAGTGGCGCGGCAACGCGGAGCGCGTACGTCCTG 289
QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
    |||
DB 290 ACCGCTCGTTACGACAGCGCCCGCGCACGAGCGGACGCGCCCTCGGTGGAGC 349
QY 101 ValAlaTyrPlyAsnAsnTyrArgAsnAlaHisSerAlaThrThrTyrSerGlyGlnTyr 120
    |||
DB 350 GTGGCTCGAAGATTAACCTACCGCAACGCCCACTCCGCGACACGTCGAGCGCGCAGTAC 409
QY 121 ValGlyGlyAlaGlyAlaArgIleAsnThrGlnTyrLeuLeuThrSerGlyThrThrGlu 140
    |||
DB 410 GTGGGGGGCGCGAGCGGAGATCAACACCGAGTGTCTCTGCTCCGCGACCAACCGAG 469
QY 141 AlaAsnAlaTyrPlySerThrLeuValGlyHisAspThrPheThrLysValLysProSer 160
    |||
DB 470 GCCAAGCCTGGAAGTCCAGCGTGTGGCCACGACCTTCACCAAGGTGAAGCGCTCC 529
QY 161 AlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsnGlyAsnProLeuAspAla 180
    |||
DB 530 GCCGCTTCATCGACGCGCGAAGAGCGCGGTCAACACGCGAACCCGCTCGACGCC 589
QY 181 ValGlnGln 183
    |||
DB 590 GTTCAGCAG 598
```

RESULT 2

US-10-013-173-1
; Sequence 1, Application US/10013173
; Publication No. US20030095977A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen C.
; APPLICANT: Graves, Scott Stoll
; APPLICANT: Schultz, Joanne Elaine
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James A.
; APPLICANT: Reno, John M.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 690022.547C1
; CURRENT APPLICATION NUMBER: US/10/013.173
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 638
; TYPE: DNA
; ORGANISM: Streptomyces avidinii
US-10-013-173-1

Alignment Scores:

Pred. No.: 1,92e-101 Length: 638
Score: 936.00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-589-870B-2 (1-183) x US-10-013-173-1 (1-638)

```
QY 1 MetArgLysIleValAlaAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
    |||
DB 50 ATGCGAAGATCGTCGTCGAGCCATCGCGTTCCCTGACCAACGCGTCTCGATTACGGCC 109
QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGlnAgi 40
    |||
DB 110 AGCGCTTCGGCAGACCCCTCCAGAGACTCGAAGGCCAGGTCTCGGCCCGCCAGGCCGC 169
QY 41 IleThrGlyThrTyrTyrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAla 60
    |||
DB 170 ATCACCAGGACCTGTTACACACAGCTCGGCTCGACTTCATCGTACCGCGGCGCGAC 229
QY 61 GlyAlaLeuThrGlyThrTyrGlySerAlaValGlyAsnAlaGlySerArgTyrValLeu 80
    |||
DB 230 GCGGCCCTGACCGAACCCTACGAGTGGCGCGGCAACGCGGAGCGCGTACGTCCTG 289
QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
    |||
DB 290 ACCGCTCGTTACGACAGCGCCCGCGCACGAGCGGACGCGCCCTCGGTGGAGC 349
QY 101 ValAlaTyrPlyAsnAsnTyrArgAsnAlaHisSerAlaThrThrTyrSerGlyGlnTyr 120
    |||
DB 350 GTGGCTCGAAGATTAACCTACCGCAACGCCCACTCCGCGACACGTCGAGCGCGCAGTAC 409
QY 121 ValGlyGlyAlaGlyAlaArgIleAsnThrGlnTyrLeuLeuThrSerGlyThrThrGlu 140
    |||
DB 410 GTGGGGGGCGCGAGCGGAGATCAACACCGAGTGTGTGCTCGACTCCGCGACCAACCGAG 469
QY 141 AlaAsnAlaTyrPlySerThrLeuValGlyHisAspThrPheThrLysValLysProSer 160
    |||
DB 470 GCCAAGCCTGGAAGTCCAGCGTGTGGCCACGACCTTCACCAAGGTGAAGCGCTCC 529
QY 161 AlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsnGlyAsnProLeuAspAla 180
    |||
DB 530 GCCGCTTCATCGACGCGCGAAGAGCGCGGTCAACACGCGAACCCGCTCGACGCC 589
QY 181 ValGlnGln 183
    |||
DB 590 GTTCAGCAG 598
```

RESULT 3

US-10-150-762-1
; Sequence 1, Application US/10150762
; Publication No. US20030103948A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen C.
; APPLICANT: Graves, Scott S.
; APPLICANT: Schultz, Joanne E.
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James A.
; APPLICANT: Reno, John M.
; APPLICANT: Deaslyne, Erica A.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; TITLE OF INVENTION: METHODS OF USE THEREOF
; FILE REFERENCE: 690022.547C2
; CURRENT APPLICATION NUMBER: US/10/150.762
; CURRENT FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 638
; TYPE: DNA


```
; ORGANISM: Streptomyces avidinii
; US-10-150-762-1

Alignment Scores:
Pred. No.: 1,92e-101      Length: 638
Score: 936.00             Matches: 183
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 14                    Gaps: 0

US-09-589-870b-2 (1-183) x US-10-150-762-1 (1-638)

QY 1 MetArgLyValIleValAlaAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
DB 50 ATCGCGAAGATCGTCGTGACGACCATCGCCGTTTCCCTGACCAAGGCTCGATTACGCC 109
QY 21 SerAlaSerAlaAspProSerLyAspSerLyAlaGlnValSerAlaAlaGlnAlaGly 40
DB 110 AGCGCTTCGCGACAGCCCTCCAAAGACTCGAAGGCCAGGCTCGGCCGCGAGCGCGC 169
QY 41 IleThrGlyThrTrpTrpYrAsnGlnLeuGlySerThrHeIleValThrAlaGlyAlaAsp 60
DB 170 ATACCGCGACCTGTGTAACAACAGCTCGCTCACTTCATCGTACCGCGCGCGCGAC 229
QY 61 GlyAlaLeuThrGlyThrTrpYrGluSerAlaValGlyAsnAlaGlnSerArgTrpValIleu 80
DB 230 GCGGCGCTGACCGGACCTTCAGAGTCCGCGCTCGCGACACCGCGCTACGTCCTG 289
QY 81 ThrGlyYrGlyYrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
DB 290 ACCGCTGCTTACGACAGCGCGCCCGCCACGACGCGACGCGCACCGCCCTCGGTGGAGC 349
QY 101 ValAlaIleTrpYrAsnAsnTrpYrArgAsnAlaHisSerAlaThrTrpSerGlyGlnTrp 120
DB 350 GTGGCTCGAAGATAACTACCGCAAGCCCACTCCCGGACCACTGGAGCGCGCGCTAC 409
QY 121 ValGlyGlyValaGlnAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrTrpGlu 140
DB 410 GTGGCGCGCGCCGAGGCGAGATCAACACCACTGGCTGCTGCTCGGACCAACCGAG 469
QY 141 AlaAsnAlaIleTrpYrSerThrLeuValGlyHisAspThrHeThrLyValLyProSer 160
DB 470 GCCAAGCGCTGGAAGTCCACGCTGTGCGCACGACCTTCCCAAGGTGAAGCGCTGC 529
QY 161 AlaAlaSerIleAspAlaAlaIleValGlyValaAsnGlnYrAsnProLeuAspAla 180
DB 530 GCGGCTTCATCCACGCGGGAAGAGCGCGCTCAACCAAGCGCAACCGCTCGACGCC 589
QY 181 ValGlnGln 183
DB 590 GTTCAGCAG 598

RESULT 4
US-10-244-821-5
; Sequence 5, Application US/10244821
; Publication NO. US20030143233A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen Charles
; APPLICANT: Graves, Scott Stoll
; APPLICANT: Schultz, Joanne Elaine
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James Allen
; APPLICANT: Reno, John M.
; APPLICANT: Dearsteyne, Erica A.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; FILE REFERENCE: 690022.547C3
; CURRENT APPLICATION NUMBER: US/10/244,821
; CURRENT FILING DATE: 2002-09-16
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
```

```
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: B9E9 single chain antibody-genomic streptavidin fusion
; US-10-244-821-5

Alignment Scores:
Pred. No.: 1,53e-91      Length: 1239
Score: 855.50           Matches: 165
Percent Similarity: 98.83%  Conservative: 4
Best Local Similarity: 96.49%  Mismatches: 1
Query Match: 91.40%      Indels: 1
DB: 12                  Gaps: 1

US-09-589-870b-2 (1-183) x US-10-244-821-5 (1-1239)

QY 14 ThrThrValSerIleThrAla---SerAlaSerAlaAspProSerLyAspSerLyAla 32
DB 724 ACCAGCGTACCGGTGACGTCTGGCTGTGGTGGACAGACCCCTCCAAAGACTCGAAGGCC 783
QY 33 GlnValSerAlaAlaGlnAlaGlyIleThrGlyThrTrpYrAsnGlnLeuGlySerThr 52
DB 784 CAGGTCTCGCGCGCGCGGACCGGCAATCACCGGCACCTGTGTAACAACGAGCTCGGCTGACC 843
QY 53 PheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTrpYrGluSerAlaValGly 72
DB 844 TTCATGTGACCGGCGCGCGCGCGCGCTGACCGGACCGAACCCTACGATCGCGCTCGGC 903
QY 73 AsnAlaGluSerArgTrpValLeuThrGlyYrArgYrAspSerAlaProAlaThrAspGly 92
DB 904 AACGCGAAGACCGGCTACGTCTGACCGGCTGTTAGACAGCGCCCGCCACCGACGCC 963
QY 93 SerGlyThrAlaLeuGlyTrpThrValAlaIleTrpYrAsnAsnTrpYrArgAsnAlaHisSer 112
DB 964 AGCGGACCGCGCTCGGTGGAGACGCGGCGCTGGAAGATAACTACCGCAACCGCCACTCC 1023
QY 113 AlaThrThrTrpSerGlyGlnTrpYrValGlyValaGlnAlaArgIleAsnThrGlnTrp 132
DB 1024 GGACCACTGTGAGCGCGCGCAATGCTGGCGGCGCGGAGCGGAGTCAACACCGCACTGG 1083
QY 133 LeuLeuThrSerGlyThrThrGlnAlaAsnAlaIleTrpYrSerThrLeuValGlyHisAsp 152
DB 1084 CTGCTACCTCCGGGACCAACGAGGCCAAGCGCTGGAATCCACGCTGTCGCGCCACGAC 1143
QY 153 ThrPheThrLyValIleProSerAlaAlaSerIleAspAlaAlaIleValaGlyVal 172
DB 1144 ACCTTCACCAAGGTGAAGCGTCCGCGCTCATGACGCGGCGGAAGAGCGCGCGCTC 1203
QY 173 AsnAsnGlyAsnProLeuAspAlaValaGlnGln 183
DB 1204 AACCAAGCAACCCGCTCGACGCGCTTCAGCAG 1236

RESULT 5
US-10-013-173-5
; Sequence 5, Application US/10013173
; Publication NO. US20030095977A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen C.
; APPLICANT: Graves, Scott Stoll
; APPLICANT: Schultz, Joanne Elaine
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James A.
; APPLICANT: Reno, John M.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; FILE REFERENCE: 690022.547C1
; CURRENT APPLICATION NUMBER: US/10/013,173
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1239
```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: B9E9 single chain antibody-genomic streptavidin fusion
US-10-013-173-5
```

```

Alignment Scores:
Pred. No.: 1,53e-91 Length: 1239
Score: 855.50 Matches: 165
Percent Similarity: 98.83% Conservative: 4
Best Local Similarity: 96.49% Mismatches: 1
Query Match: 91.40% Indels: 1
DB: 14 Gaps: 1
```

US-09-589-870B-2 (1-183) x US-10-013-173-5 (1-1239)

```

QY 14 ThrThrValSerIleThrAla---SerAlaSerAlaAspProSerIleAspSerIleValAla 32
DB 724 ACCACGGTCACCGTAGCTCTGGCTTGGTGGACGACCCCTCCAGAGACTCGAAGGCC 783
QY 33 GlnValSerAlaAlaGlnAlaGlyIleThrGlyThrTrpTyraSngInLeuGlySerThr 52
DB 784 CAGGCTCGGCGCGCGAGCGCCGACCTGATCAACCGGACCTGTTACAACAGCTCGGCTCGACC 843
QY 53 PheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyrgluSerAlaValAlgly 72
DB 844 TTCATCGTAGACCGCGGCGCGACCGCCCTGACCGAGAACTACGAGTCGGCGCTCGGC 903
QY 73 AsnAlaGluSerArgTyValleuThrglyArgTyraSpsSerAlaProAlaThrAspGly 92
DB 904 AACGCCGAGAGCGGTACCTCTGACGGATCTGACGACGAGCGCCCGCGCCAGCCGACGCGC 963
QY 93 SerGlyThrAlaLeuGlyTrpThrValAlaTrpIleAsnTyraSngInLeuValHisSer 112
DB 964 AGCGGACCGCCCTCGTGGACGGTGGCTGGAAGAAATTAATCAACCGACCCCACTCC 1023
QY 113 AlaThrThrTrpSerGlyGlnTyValAlglyAlaGlnAlaArgIleAsnThrGlnTrp 132
DB 1024 GCGACACAGTGAGCGGCGCAGTACGTCGCGCGCGCGAGCGAGATCAACACCCAGTGG 1083
QY 133 LeuLeuThrSerGlyThrThrglnAlaAsnAlaTrpIleSerThrIleValGlyHisAsp 152
DB 1084 CTGCTGACCTCGGACCAACCGAGCGCAAGCTTGAACTCCACGCTGTCGCGCACGAC 1143
QY 153 ThrPheThrIleValIleProSerAlaAlaSerIleAspAlaAlaIleValAlglyVal 172
DB 1144 ACCTTCACCAAGGTGAAGCCGTCGCGCTCCATCGACGCGCGGAGAAAGCCGCGCTC 1203
QY 173 AsnAsnGlyAsnProLeuAspAlaValGlnGln 183
DB 1204 AACCAACGGCAACCCGCTCGACGCCGCTTCAGCAG 1236
```

RESULT 6

```

US-10-150-762-5
; Sequence 5, Application US/10150762
; Publication No. US20030103948A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen C.
; APPLICANT: Graves, Scott S.
; APPLICANT: Schultz, Joanne E.
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James A.
; APPLICANT: Reno, John M.
; APPLICANT: Deastyne, Erica A.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; FILE REFERENCE: 690022.547C2
; CURRENT APPLICATION NUMBER: US/10/150,762
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1239
```

```

; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: B9E9 single chain antibody-genomic streptavidin fusion
US-10-150-762-5
```

```

Alignment Scores:
Pred. No.: 1,53e-91 Length: 1239
Score: 855.50 Matches: 165
Percent Similarity: 98.83% Conservative: 4
Best Local Similarity: 96.49% Mismatches: 1
Query Match: 91.40% Indels: 1
DB: 14 Gaps: 1
```

US-09-589-870B-2 (1-183) x US-10-150-762-5 (1-1239)

```

QY 14 ThrThrValSerIleThrAla---SerAlaSerAlaAspProSerIleAspSerIleValAla 32
DB 724 ACCACGGTCACCGTAGCTCTGGCTTGGTGGACGACCCCTCCAGAGACTCGAAGGCC 783
QY 33 GlnValSerAlaAlaGlnAlaGlyIleThrGlyThrTrpTyraSngInLeuGlySerThr 52
DB 784 CAGGCTCGGCGCGCGAGCGCCGACCTGATCAACCGGACCTGTTACAACAGCTCGGCTCGACC 843
QY 53 PheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyrgluSerAlaValAlgly 72
DB 844 TTCATCGTAGACCGCGGCGCGACCGCCCTGACCGAGAACTACGAGTCGGCGCTCGGC 903
QY 73 AsnAlaGluSerArgTyValleuThrglyArgTyraSpsSerAlaProAlaThrAspGly 92
DB 904 AACGCCGAGAGCGGTACCTCTGACGGATCTGACGACGAGCGCCCGCGCCAGCCGACGCGC 963
QY 93 SerGlyThrAlaLeuGlyTrpThrValAlaTrpIleAsnTyraSngInLeuValHisSer 112
DB 964 AGCGGACCGCCCTCGTGGACGGTGGCTGGAAGAAATTAATCAACCGACCCCACTCC 1023
QY 113 AlaThrThrTrpSerGlyGlnTyValAlglyAlaGlnAlaArgIleAsnThrGlnTrp 132
DB 1024 GCGACACAGTGAGCGGCGCAGTACCTCGCGCGCGCGAGCGAGATCAACACCCAGTGG 1083
QY 133 LeuLeuThrSerGlyThrThrglnAlaAsnAlaTrpIleSerThrIleValGlyHisAsp 152
DB 1084 CTGCTGACCTCGGACCAACCGAGCGCAAGCTTGAACTCCACGCTGTCGCGCACGAC 1143
QY 153 ThrPheThrIleValIleProSerAlaAlaSerIleAspAlaAlaIleValAlglyVal 172
DB 1144 ACCTTCACCAAGGTGAAGCCGTCGCGCTCCATCGACGCGCGGAGAAAGCCGCGCTC 1203
QY 173 AsnAsnGlyAsnProLeuAspAlaValGlnGln 183
DB 1204 AACCAACGGCAACCCGCTCGACGCCGCTTCAGCAG 1236
```

RESULT 7

```

US-10-244-821-3
; Sequence 3, Application US/10244821
; Publication No. US2003014323A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen Charles
; APPLICANT: Graves, Scott Stoll
; APPLICANT: Schultz, Joanne Elaine
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James Allen
; APPLICANT: Reno, John M.
; APPLICANT: Deastyne, Erica A.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; FILE REFERENCE: 690022.547C3
; CURRENT APPLICATION NUMBER: US/10/244,821
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1614
```

```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: hunR-LU-10 single chain antibody-genomic streptavidin
; US-10-244-821-3

Alignment Scores:
Pred. No.: 2,84e-91 Length: 1614
Score: 854.50 Matches: 166
Percent Similarity: 97.67% Conservative: 2
Best Local Similarity: 96.51% Mismatches: 3
Query Match: 91.29% Indels: 1
DB: 12 Gaps: 1

US-09-589-870b-2 (1-183) x US-10-244-821-3 (1-1614)

QY 12 SerLeuThrThrValSerIleThrAlaSerAlaAspProSerLyAspSerLyS 31
Db 1093 ACCTTAGTCACCGTGAGC---TCTGGCTGTGGTTCGGCAGACCCCTCCAAAGACTCGAAG 1149

QY 32 AlaGlnValSerAlaAlaGlnAlaGlyIleThrGlyThrTyrAsnGlnLeuGlySer 51
Db 1150 GCCCAGGCTCGGCCCGCCGAGCGGCATCACCGGCACTGTGTACAAACGACTCGGCTCG 1209

QY 52 ThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyrGluSerAlaVal 71
Db 1210 ACCTTCATCTGTACCGCGCGCCGCGCGCCCTGACCGGAACCTTACAGATCGGCCGTC 1269

QY 72 GlyAsnAlaGluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAsp 91
Db 1270 GGCAACGCCGAGAGCGCGCTACGTCCTGACCGGTCGTAGACACGCCCGCGCCGACCGAC 1329

QY 92 GlySerGlyThrAlaLeuGlyTyrThrValAlaTyrPlyAsnAsnTyrArgAsnAlaHis 111
Db 1330 GGCAGCGGACCGCCCTCGGTGGACGCTGGCAAGAAATTAACCGCAACGCCCGCAC 1389

QY 112 SerAlaThrThrTyrSerGlyGlnTyrValGlyAlaGlnAlaArgIleAsnThrGln 131
Db 1390 TCCGGACCAACGTGAGCGCGCCAGTACGTCCGGCGCGCGAGCGAGATCAACACCCGAC 1449

QY 132 TrpLeuLeuThrSerGlyThrThrGlnAlaAsnAlaTyrPlySerThrLeuValGlyHis 151
Db 1450 TGGCTGCTGACCTCCGGACCAACCGAGCGCAACGCTGGAAGTCCACGCTGGTCGGCCAC 1509

QY 152 AspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAlaLysLysAlaGly 171
Db 1510 GACACCTTCACCAAGGTGAAGCGTCCGCCGCTTCATCGACGCGCGAAGAGCGCGGC 1569

QY 172 ValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
Db 1570 GTCAACAAAGCGCAACCCGCTCGAGCGCTTCAGCAG 1605

RESULT 8
US-10-013-173-3
; Sequence 3, Application US/10013173
; Publication No. US20030095977A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen C.
; APPLICANT: Graves, Scott Stoll
; APPLICANT: Schultz, Joanne Elaine
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James A.
; APPLICANT: Reno, John M.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; FILE REFERENCE: 690022.547C1
; CURRENT APPLICATION NUMBER: US/10/013,173
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1614
```

```
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: hunR-LU-10 single chain antibody-genomic streptavidin
; US-10-013-173-3

Alignment Scores:
Pred. No.: 2,84e-91 Length: 1614
Score: 854.50 Matches: 166
Percent Similarity: 97.67% Conservative: 2
Best Local Similarity: 96.51% Mismatches: 3
Query Match: 91.29% Indels: 1
DB: 14 Gaps: 1

US-09-589-870b-2 (1-183) x US-10-013-173-3 (1-1614)

QY 12 SerLeuThrThrValSerIleThrAlaSerAlaAspProSerLyAspSerLyS 31
Db 1093 ACCTTAGTCACCGTGAGC---TCTGGCTGTGGTTCGGCAGACCCCTCCAAAGACTCGAAG 1149

QY 32 AlaGlnValSerAlaAlaGlnAlaGlyIleThrGlyThrTyrAsnGlnLeuGlySer 51
Db 1150 GCCCAGGCTCGGCCCGCCGAGCGGCATCACCGGCACTGTGTACAAACGACTCGGCTCG 1209

QY 52 ThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyrGluSerAlaVal 71
Db 1210 ACCTTCATCTGTACCGCGCGCCGCGCGCCCTGACCGGAACCTTACAGATCGGCCGTC 1269

QY 72 GlyAsnAlaGluSerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAsp 91
Db 1270 GGCAACGCCGAGAGCGCGCTACGTCCTGACCGGTCGTAGACACGCCCGCGCCGACCGAC 1329

QY 92 GlySerGlyThrAlaLeuGlyTyrThrValAlaTyrPlyAsnAsnTyrArgAsnAlaHis 111
Db 1330 GGCAGCGGACCGCCCTCGGTGGACGCTGGCAAGAAATTAACCGCAACGCCCGCAC 1389

QY 112 SerAlaThrThrTyrSerGlyGlnTyrValGlyAlaGlnAlaArgIleAsnThrGln 131
Db 1390 TCCGGACCAACGTGAGCGCGCCAGTACGTCCGGCGCGCGAGCGAGATCAACACCCGAC 1449

QY 132 TrpLeuLeuThrSerGlyThrThrGlnAlaAsnAlaTyrPlySerThrLeuValGlyHis 151
Db 1450 TGGCTGCTGACCTCCGGACCAACCGAGCGCAACGCTGGAAGTCCACGCTGGTCGGCCAC 1509

QY 152 AspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAlaLysLysAlaGly 171
Db 1510 GACACCTTCACCAAGGTGAAGCGTCCGCCGCTTCATCGACGCGCGAAGAGCGCGGC 1569

QY 172 ValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
Db 1570 GTCAACAAAGCGCAACCCGCTCGAGCGCTTCAGCAG 1605

RESULT 9
US-10-150-762-3
; Sequence 3, Application US/10150762
; Publication No. US20030103948A1
; GENERAL INFORMATION:
; APPLICANT: Goshorn, Stephen C.
; APPLICANT: Graves, Scott S.
; APPLICANT: Schultz, Joanne E.
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James A.
; APPLICANT: Reno, John M.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; FILE REFERENCE: 690022.547C2
; CURRENT APPLICATION NUMBER: US/10/150,762
; CURRENT FILING DATE: 2002-05-17
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
```

```

; LENGTH: 1614
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: humr-LU-10 single chain antibody-genomic streptavidin
US-10-150-762-3

```

Alignment Scores:

Pred. No.:	2,84e-91	Length:	1614
Score:	854.50	Matches:	166
Percent Similarity:	97.67%	Conservative:	2
Best Local Similarity:	96.51%	Mismatches:	3
Query Match:	91.29%	Indels:	1
DB:	14	Gaps:	1

US-09-589-870b-2 (1-183) x US-10-150-762-3 (1-1614)

```

QY 12 SerLeuThrThrValSerIleThrAlaSerAlaSerAlaAspProSerLysAspSerLys 31
   :::::
Db 1093 ACCTTAGTCCAGCCGCTGAGC---TCTGGCTCTGCTTCCGCGACACCCCTCCAAAGACTCGAAG 1149

QY 32 AlaGlnValSerAlaAlaGlnValGlyIleThrGlyThrTrpTyraSngInleuGlySer 51
   |||||
Db 1150 GCCGAGGTTCCGCGCCGCGGCGGCGCATCCCGGACCTGGTACACAGCTCGGCTCG 1209

QY 52 ThrPheIleValThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyrgIuSerAlaVal 71
   |||||
Db 1210 ACCTTCATCGTACCGCGCGCGCGCCGACCGGCTTACCGAAGCTACGAGTCCGCGCTAC 1269

QY 72 GlyAsnAlaGluSerArgTyValLeuThrGlyValArgTyAspSerAlaProAlaThrAsp 91
   |||||
Db 1270 GCCAAGCGCGAAGCGCGCTACGTCCTGACCGGCTGTTACGACAGCGCCCGCGCACCGAC 1329

QY 92 GlySerGlyThrAlaLeuGlyTrpThrValAlaTrpLysAsnAsnTyraSngInleuHis 111
   |||||
Db 1330 GCGAGGGGACCGCCCTCGGTTGGAGGCGGCGCTGGAAATTAACCTACCGCAACGCCAC 1389

QY 112 SerAlaThrThrTrpSerGlyGlnTyValGlyAlaGlyAlaArgIleAsnThrGln 131
   |||||
Db 1390 TCCGGGACCGACGTGGAGCGCGGACGTACGTCGCGCGCGGCGGAGCGAGATCAACACCCAG 1449

QY 132 TrpLeuLeuThrSerGlyThrThrGluAlaAsnAlaTrpLysSerThrIleuValGlyHis 151
   |||||
Db 1450 TGGCTGCTGACCTCCGCGCACACCGGCGCAACGCTGGAGTCCACGCTGCTGCGCGCAC 1509

QY 152 AspThrPheThrLysValLysProSerAlaAlaSerIleAspAlaAlaLysLysAlaGly 171
   |||||
Db 1510 GACACCTTCACCAAGGTGAAGCGTCCGCGCTCCATCGACGCGCGAAGAGCGCGCGC 1569

QY 172 ValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
   |||||
Db 1570 GTCAACAACGCGCACCCGCTGACGCGCTTCAGCAG 1605

```

RESULT 10
US/10/244
Sequence 7, Application US/10244821
Publication No. US2003014323A1
GENERAL INFORMATION:

```

; APPLICANT: Goshorn, Stephen Charles
; APPLICANT: Graves, Scott Stoll
; APPLICANT: Schultz, Joanne Elaine
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James Allen
; APPLICANT: Reno, John M.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; FILE REFERENCE: 690022.547C3
; CURRENT APPLICATION NUMBER: US/10/244,821
; CURRENT FILING DATE: 2002-09-16
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 7
; LENGTH: 1280
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: B9E9 single chain antibody-genomic streptavidin fusion construct
US/10/244,821-7

```

Alignment Scores:

Pred. No.:	1.61e-90	Length:	1280
Score:	847.00 <td>Matches:</td> <td>162 </td>	Matches:	162
Percent Similarity:	98.79% <td>Conservative:</td> <td>1 </td>	Conservative:	1
Best Local Similarity:	98.18% <td>Mismatches:</td> <td>2 </td>	Mismatches:	2
Query Match:	90.49% <td>Indels:</td> <td>0 </td>	Indels:	0
DB:	12	Gaps:	0

US-09-589-870b-2 (1-183) x US/10/244 (1-1280)

```

QY 19 ThrAlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGln 38
   :::::
Db 777 TCTGGCTCTGCTTCCGCGACACCCCTCCAAAGACTCGAAGCCCAAGTCTCGGCGCGCGAG 836

QY 39 AlaGlyIleThrGlyThrTrpTyraSngInleuGlySerThrPheIleValThrAlaGly 58
   |||||
Db 837 GCCGCAATCCCGGCGACCTGGTACACCAAGCTCGGCTGACCTTCATCTGACCGCGGCGC 896

QY 59 AlaAspGlyAlaLeuThrGlyThrTyrgIuSerAlaValGlyAsnAlaGluSerArgTyr 78
   |||||
Db 897 GCCGAGCGCGCCCTTACCGGAACTACGAGTCCGCGCTGGCAACGCCGAGACGCCGCTAC 956

QY 79 ValLeuThrGlyValArgTyAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGly 98
   |||||
Db 957 GTCCGACCGGCTGTTACGACAGCGCCCGCGCACCGGCGAGCGGACCGCGCTCCGT 1016

QY 99 TrpThrValAlaTrpLysAsnAsnTyraSngInleuHisSerAlaThrThrTrpSerGly 118
   |||||
Db 1017 TGGAGGCGGCTGGAAATTAACCTACCGCAACGCCACTCGGACCAACGATGGAGCGCGC 1076

QY 119 GlnTyValGlyGlyAlaGlyAlaArgIleAsnThrGlnTrpLeuThrSerGlyThr 138
   |||||
Db 1077 CAGTACGTCGCGCGCGCGCGGCGGAGCGAGATCAACACCGAGTGTGCTGACCTCCGCGCAC 1136

QY 139 ThrGluAlaAsnAlaTrpLysSerThrIleuValGlyHisAspThrPheThrLysValLys 158
   |||||
Db 1137 ACCGAGCGCAACGCGCTGAAATTCACGCTGCTGCGCGCACACCTTCACCAAGGTGAAG 1196

QY 159 ProSerAlaAlaSerIleAspAlaAlaLysLysAlaGlyValaAsnGlyAsnProLeu 178
   |||||
Db 1197 CGTCCGCGCGCTCCATCGACGCGCGGCGAAGAGCGGCGTCAACACGCGACCCGCTC 1256

QY 179 AspAlaValGlnGln 183
   |||||
Db 1257 GACGCGCTTCAGCAG 1271

```

RESULT 11
US/10/013
Sequence 7, Application US/10013173
Publication No. US20030095977A1
GENERAL INFORMATION:

```

; APPLICANT: Goshorn, Stephen C.
; APPLICANT: Graves, Scott Stoll
; APPLICANT: Schultz, Joanne Elaine
; APPLICANT: Lin, Yukang
; APPLICANT: Sanderson, James A.
; APPLICANT: Reno, John M.
; TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
; FILE REFERENCE: 690022.547C1
; CURRENT APPLICATION NUMBER: US/10/013,173
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7

```


LENGTH: 1371
 TYPE: DNA
 ORGANISM: Mus musculus
 US-10-244-821-87

Alignment Scores:

Pred. No.:	1.76e-90	Length:	1371
Score:	847.00	Matches:	162
Percent Similarity:	98.79%	Conservative:	1
Best Local Similarity:	98.18%	Mismatches:	2
Query Match:	90.49%	Indels:	0
DB:	12	Gaps:	0

US-09-589-870B-2 (1-183) x US-10-244-821-87 (1-1371)

```

QY 19 ThrAlaSerAlaSerAlaAspProSerLyAspSerLyAlaGlnValSerAlaAlaGlu 38
Db 874 TCTGGCTGTGTTGGCAGACCCCTCCAGAGCTCGAAGGCCAGGTCTCGGCCGCGAG 933
QY 39 AlaGlyIleThrGlyThrTyrAsnGlnLeuGlySerThrPheIleValThrAlaGly 58
Db 934 GCCGGCATCACCGGACCTGTAACAACAGCTCGCTCGACTTCATCTGTACCGCGGCG 993
QY 59 AlaAspGlyAlaLeuThrGlyThrTyrGluSerAlaValAlaGlyAsnAlaGluSerArgTyr 78
Db 994 GCCGAGCGGCGCCCTGACCGGAACCTACGAGTCGGCGCTCGGCAACGCCGAGCGGCTAC 1053
QY 79 ValLeuThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGly 98
Db 1054 GTCTGACCGGTGCTTACACAGCGCCCGGCAACGAGCGGAGCGGCGCCCTCGGT 1113
QY 99 TrpThrValAlaIleTyrLyAsnAsnTyrArgAsnAlaHisSerAlaThrThrTyrSerGly 118
Db 1114 TGGAGCGGTGCTGGAAGAATACATACCGCAACGCCACTCCGCGGACCACTGTGAGCGGCG 1173
QY 119 GlnTyrValAlaGlyAlaGluAlaAlaArgIleAsnThrGlnTyrLeuLeuThrSerGlyThr 138
Db 1174 CAGTACGTGCGGCGGCGCCGAGGAGATCAACACCACTGCTGTGACCTCCGCGCAC 1233
QY 139 ThrGluAlaAsnAlaTyrLySerThrLeuValGlyHisAspThrPheThrIlyValIys 158
Db 1234 ACCGAGGCCAAGCCTGGAAGTCCAGCTGTGCGGCAAGACCTTCCCAAGGTGAAG 1293
QY 159 ProSerAlaAlaSerIleAspAlaAlaIlyslYsAlaGlyValAsnAsnGlyAsnProLeu 178
Db 1294 CCGTCGCGCGCTTCATCGACGCGGGAAGAGCGGCGTCAACACGCAACCGCGCTC 1353
QY 179 AspAlaValAlaGlnGln 183
Db 1354 GACGCCGTTCAAGCAG 1368

```

RESULT 14

US-10-244-821-48
 Sequence 48, Application US/10244821
 Publication No. US2003014323A1
 GENERAL INFORMATION:
 APPLICANT: Goshorn, Stephen Charles
 APPLICANT: Graves, Scott Stoll
 APPLICANT: Schultz, Joanne Elaine
 APPLICANT: Lin, Yukang
 APPLICANT: Sanderson, James Allen
 APPLICANT: Reno, John M.
 TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
 TITLE OF INVENTION: METHODS OF USE THEREOF
 FILE REFERENCE: 690022.547C3
 CURRENT APPLICATION NUMBER: US/10/244,821
 CURRENT FILING DATE: 2002-09-16
 NUMBER OF SEQ ID NOS: 92
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 48
 LENGTH: 1467
 TYPE: DNA

ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: CC49 single chain antibody-*genomic streptavidin*
 OTHER INFORMATION: fusion sequence
 US-10-244-821-48

Alignment Scores:

Pred. No.:	1.93e-90	Length:	1467
Score:	847.00	Matches:	162
Percent Similarity:	98.79%	Conservative:	1
Best Local Similarity:	98.18%	Mismatches:	2
Query Match:	90.49%	Indels:	0
DB:	12	Gaps:	0

US-09-589-870B-2 (1-183) x US-10-244-821-48 (1-1467)

```

QY 19 ThrAlaSerAlaSerAlaAspProSerLyAspSerLyAlaGlnValSerAlaAlaGlu 38
Db 898 TCTGGCTGTGTTGGCAGACCCCTCCAGAGCTCGAAGGCCAGGTCTCGGCCGCGAG 957
QY 39 AlaGlyIleThrGlyThrTyrAsnGlnLeuGlySerThrPheIleValThrAlaGly 58
Db 958 GCCGCATCACCGGACCTGTAACAACAGCTCGCTCGACTTCATCTGTACCGCGGCG 1017
QY 59 AlaAspGlyAlaLeuThrGlyThrTyrGluSerAlaValAlaGlyAsnAlaGluSerArgTyr 78
Db 1018 GCCGAGCGGCGCCCTGACCGGAACCTACGAGTCGGCGCTCGGCAACGCCGAGCGGCTAC 1077
QY 79 ValLeuThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGly 98
Db 1078 GTCTGACCGGTGCTTACAGACGCGCCCGGCAACGAGCGGAGCGGCGCCCTCGGT 1137
QY 99 TrpThrValAlaIleTyrLyAsnAsnTyrArgAsnAlaHisSerAlaThrThrTyrSerGly 118
Db 1138 TGGAGCGGTGCTGGAAGAATACATACCGCAACGCCACTCCGCGACCACTGTGAGCGGCG 1197
QY 119 GlnTyrValAlaGlyAlaGluAlaAlaArgIleAsnThrGlnTyrLeuLeuThrSerGlyThr 138
Db 1198 CAGTACGTGCGGCGGCGGCGGAGGATCAACACCCAGTGTGCTGACCTCCGCGCAC 1257
QY 139 ThrGluAlaAsnAlaTyrLySerThrLeuValGlyHisAspThrPheThrIlyValIys 158
Db 1258 ACCGAGGCCAAGCCTGGAAGTCCAGCTGTGCGGCAAGACCTTCCCAAGGTGAAG 1317
QY 159 ProSerAlaAlaSerIleAspAlaAlaIlyslYsAlaGlyValAsnAsnGlyAsnProLeu 178
Db 1318 CCGTCGCGCGCTTCATCGACGCGGGAAGAGCGGCGTCAACACGCAACCGCGCTC 1377
QY 179 AspAlaValAlaGlnGln 183
Db 1378 GACGCCGTTCAAGCAG 1392

```

RESULT 15

US-10-013-173-48
 Sequence 48, Application US/10013173
 Publication No. US20030095977A1
 GENERAL INFORMATION:
 APPLICANT: Goshorn, Stephen C.
 APPLICANT: Graves, Scott Stoll
 APPLICANT: Schultz, Joanne Elaine
 APPLICANT: Lin, Yukang
 APPLICANT: Sanderson, James A.
 APPLICANT: Reno, John M.
 TITLE OF INVENTION: STREPTAVIDIN EXPRESSED GENE FUSIONS AND
 TITLE OF INVENTION: METHODS OF USE THEREOF
 FILE REFERENCE: 690022.547C1
 CURRENT APPLICATION NUMBER: US/10/013,173
 CURRENT FILING DATE: 2001-12-07
 NUMBER OF SEQ ID NOS: 69
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 48
 LENGTH: 1467
 TYPE: DNA

```
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CC49 single chain antibody-genomic streptavidin
; OTHER INFORMATION: fusion sequence
US-10-013-173-48
```

Alignment Scores:

Pred. No.:	1,93e-90	length:	1467
Score:	847.00	Matches:	162
Percent Similarity:	98.79%	Conservative:	1
Best Local Similarity:	98.18%	Mismatches:	2
Query Match:	90.49%	Indels:	0
DB:	14	Gaps:	0

US-09-589-870b-2 (1-183) x US-10-013-173-48 (1-1467)

```
QY 19 ThrAlaSerAlaSerAlaAppProSerLysAppSerLysAlaGlnValSerAlaAlaGlu 38
   |||
Db 898 TCTGGCTCTGCTGTCGCGAGACCCCTCCAGAGACTCGAAGGCCGAGTCTCGCCGCGAG 957

QY 39 AlaGlyIleThrGlyThrTPTPTyRAenGlnLeuGlySerThrPheIleValThrAlaGly 58
   |||
Db 958 GCCGCGATCACCGGCACCTGTGTACACACAGCTCGCTCGACTTCATCTGACCGCGGC 1017

QY 59 AlaAPGlyValaLeuThrGlyThrTyGluSerAlaValGlyValaenAlaGluSerArgTyr 78
   |||
Db 1018 GCCGACGCGCCCTCGACCGGACCTACGAGTCCGCCCTCGCAACGCCGAGACCGCTAC 1077

QY 79 ValLeuThrGlyValGlyTyRApSerAlaProAlaThrApGlySerglyThrAlaLeuGly 98
   |||
Db 1078 GTCCTGACCGGTGCTTACGACAGCGCCCGGCCGACGCGAGCGGCACCGCCCTCGGT 1137

QY 99 TrpThrValAlaTrpLysaenAntyArgAsnAlaHisSerAlaThrThrTrpSergly 118
   |||
Db 1138 TGGACGGTGGCTGGAGAAATTAATACTACCGACGCCCTCCGACACAGTGAGCGGC 1197

QY 119 GlnTyValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuLeuThrSerglyThr 138
   |||
Db 1198 CAGTACCTCGCGCGCGCGCGGAGATCAACACCACTGGCTGCTGACCTCGGCACC 1257

QY 139 ThrGluAlaAsnAlaTrpLysSerThrLeuValGlyHisApThrPheThrLysValLys 158
   |||
Db 1258 ACCGAGGCCAAGCGCTGGAAGTCCACGCTGTCGCGCACACCTTCACCAAGTGTAAG 1317

QY 159 ProSerAlaAlaSerIleAspAlaAlaLysValAlaGlyValaenGlyAsnProLeu 178
   |||
Db 1318 CCGTCCGCGCGCTCCATCGACGCGCGGAGAGAGCGCGCTCAACACGCGCAACCGCTC 1377

QY 179 AspAlaValGlnGln 183
   |||
Db 1378 GACGCCGCTCAGCAG 1392
```

Search completed: October 28, 2003, 03:00:31
Job time : 259 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: October 27, 2003, 21:58:22 : Search time 61 Seconds
(without alignments)
1324.149 Million cell updates/sec

Title: US-09-589-870B-2

Perfect score: 936

Sequence: 1 MKRIVAAIVASLTVSITA.....IDAKKAGVNGNPLDAVQO 183

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame.p2n.model -DEV=x1h
-Q=/cg2_1/USPTO.epool/US0589870/runac.27102003_104447_16637/app.query.fasta_1.327
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=opt -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0589870.@CGN_1.1.56 -runac.27102003_104447_16637 -NCPu=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLCK=100 -LONLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_NA.*
1: /cg2_6/prodata/2/ina/58_COMB.seq.*
2: /cg2_6/prodata/2/ina/58_COMB.seq.*
3: /cg2_6/prodata/2/ina/58_COMB.seq.*
4: /cg2_6/prodata/2/ina/58_COMB.seq.*
5: /cg2_6/prodata/2/ina/pctus_COMB.seq.*
6: /cg2_6/prodata/2/ina/backfillseq.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	936	100.0	552	5 PCT-US93-05240-13	Sequence 13, Appl
2	936	100.0	638	4 US-08-831-399-1	Sequence 1, Appl
3	936	100.0	638	4 US-09-381-430-1	Sequence 1, Appl
4	936	100.0	638	4 US-09-368-862-1	Sequence 1, Appl
5	936	100.0	638	4 US-09-368-772-1	Sequence 1, Appl
6	936	100.0	1131	6 5168049-1	Patent No. 5168049
7	846	90.4	1356	2 US-08-491-988-4	Sequence 4, Appl
8	837.5	89.5	525	1 US-07-924-028A-2	Sequence 2, Appl
9	745	79.6	1286	2 US-08-491-988-6	Sequence 6, Appl
10	687.5	73.5	1173	4 US-09-142-974B-4	Sequence 4, Appl
11	682	72.9	1176	4 US-09-142-974B-3	Sequence 3, Appl
12	681.5	72.8	1257	2 US-08-491-988-8	Sequence 8, Appl

13	675	72.1	384	4 US-08-831-399-15	Sequence 15, Appl
14	675	72.1	384	4 US-09-366-862-15	Sequence 15, Appl
15	675	72.1	384	4 US-09-368-772-15	Sequence 15, Appl
16	675	72.1	387	1 US-08-211-833-1	Sequence 1, Appl
17	675	72.1	387	1 US-08-434-718-1	Sequence 1, Appl
18	675	72.1	387	1 US-07-780-717C-6	Sequence 6, Appl
19	145	15.5	604	4 US-08-831-399-3	Sequence 3, Appl
20	145	15.5	604	4 US-09-366-862-3	Sequence 3, Appl
21	145	15.5	604	4 US-09-368-772-3	Sequence 3, Appl
22	142.5	15.2	484	1 US-08-554-586-1	Sequence 1, Appl
23	110.5	11.8	3381	4 US-09-336-447A-6	Sequence 6, Appl
24	105.5	11.3	3018	4 US-09-206-942-40	Sequence 40, Appl
25	105.5	11.3	3036	4 US-09-206-942-38	Sequence 38, Appl
26	105.5	11.3	4131	4 US-09-252-991A-13873	Sequence 13873, A
27	105.5	11.3	4329	4 US-09-252-991A-13873	Sequence 13873, A
28	105.5	11.3	8211	4 US-09-252-991A-13873	Sequence 13873, A
29	101.5	10.8	3300	2 US-08-194-290-6	Sequence 6, Appl
30	101.5	10.8	3300	2 US-08-194-290-6	Sequence 6, Appl
31	101.5	10.8	3300	2 US-08-194-290-6	Sequence 6, Appl
32	99.5	10.6	3108	4 US-09-142-648B-6	Sequence 6, Appl
33	99.5	10.6	4937	1 US-08-038-682-3	Sequence 72, Appl
34	99.5	10.6	4937	1 US-08-038-682-3	Sequence 3, Appl
35	99.5	10.6	4937	2 US-08-302-832-3	Sequence 3, Appl
36	99.5	10.6	4937	2 US-08-530-198-3	Sequence 3, Appl
37	99.5	10.6	4937	2 US-08-469-880-3	Sequence 3, Appl
38	99.5	10.6	4937	2 US-08-728-470-3	Sequence 3, Appl
39	99.5	10.6	4937	2 US-08-617-697-3	Sequence 3, Appl
40	99.5	10.6	4937	4 US-08-719-641-3	Sequence 3, Appl
41	99.5	10.6	4937	4 US-09-206-942-70	Sequence 70, Appl
42	99.5	10.6	9323	1 US-08-038-682-6	Sequence 6, Appl
43	99.5	10.6	9323	1 US-08-302-832-6	Sequence 6, Appl
44	99.5	10.6	9323	2 US-08-530-198-6	Sequence 6, Appl
45	99.5	10.6	9323	2 US-08-469-880-6	Sequence 6, Appl
				2 US-08-728-470-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1
PCT-US93-05240-13
Sequence 13, Application PC/TUS9305240
GENERAL INFORMATION:
APPLICANT: NAGARAJAN, VASANTHA
TITLE OF INVENTION: PRODUCTION OF STREPTAVIDIN FROM BACILLUS
TITLE OF INVENTION: SUBTILLIS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: DU PONT COMPANY
STREET: BARLEY MILL PLAZA 36
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: USA
ZIP: 19880-0036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/05240
FILING DATE: 19930527
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: GEIGER, KATHLEEN W
REFERENCE/DOCKET NUMBER: CR 9029
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-1118
TELEFAX: 302-892-7949
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 552 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
PCT-US93-05240-13

Alignment Scores:
Pred. No.: 3,02e-91 Length: 552
Score: 936.00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 5 Gaps: 0

US-09-589-870B-2 (1-183) x PCT-US93-05240-13 (1-552)

```
QY 1 MetArgLysIleValAlaAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
DB 1 ATGCGCAGATCGTGTGACACCTTCGCGGTTTCCCTGACACCGGTCTGATACGGCC 60
QY 21 SerAlaSerAlaAspProSerIleAspSerIleValSerAlaAlaGluAlaGly 40
DB 61 AGCGCTTGGCAGACCTCCCAAGACTCGAAGGCCCAAGTCTCGCGCGCCAGGCCGAC 120
QY 41 IleThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAlaAsp 60
DB 121 ATCACCAGCAGCTGTGACACAGCTCGGCTGACCTTCTGATCGGCGGCCGCGAC 180
QY 61 GlyAlaLeuThrGlyThrTyrGluSerAlaValGlyAsnAlaGluSerArgTyrValLeu 80
DB 181 GGCGCCCTGACCGGAACCTACAGAGTCGGCCGCGCAAGCCGAGAGCGGCTACGCTCTG 240
QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
DB 241 ACCGGTGTGAGCAGAGCGCCCGCCAGCCAGCGAGCGGACCGCCCTCGGTGAGC 300
QY 101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyr 120
DB 301 GTGGCTTGGAGAAATTAACCTACCGCAACGCCCACTCCGACACCACTGAGCGGCCAGTTC 360
QY 121 ValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuThrSerGlyThrThrGlu 140
DB 361 GTCCGCGGCGCGCAGCGGATCAACACCGAGTGGCTGCTGACCTCCGCGGACCGCAG 420
QY 141 AlaAsnAlaTrpLysSerThrIleuValGlyHisAspThrPheThrLysValLysProSer 160
DB 421 GCCAAGCGCTGGAATTCACAGCTGTGCGCCACGACACCTTACCAAGGTGAAGCGCTCC 480
QY 161 AlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsnGlyAsnProLeuAspAla 180
DB 481 GCCGCTTCATGACGCGCGGAGAAAGCGCGCTCAACACAGCGCAACCCGCTCGACGCC 540
QY 181 ValGlnGln 183
DB 541 GTTCAGCAG 549
```

RESULT 2

US-08-831-399-1
Sequence 1, Application US/08811399
Patent No. 6312916

GENERAL INFORMATION:

APPLICANT: Kopeckzki, Erhard; Muller, Rainer;
APPLICANT: Engh, Richard; Schmitt, Urban; Deger, Arno; Brandstetter, Hans
TITLE OF INVENTION: Recombinant Inactive Core
TITLE OF INVENTION: Streptavidin Mutants
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
COUNTRY: USA
ZIP: 10022

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/831,399
FILING DATE: 1-April-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 196 13 053.0
FILING DATE: 1-April-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE 196 37 718.8
FILING DATE: 16-September-1996
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6312916man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: HUBR 1105
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 638 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 50..598
OTHER INFORMATION:
OTHER INFORMATION: peptide, and 122..598 to mat peptide.

US-08-831-399-1

Alignment Scores:
Pred. No.: 3.62e-91 Length: 638
Score: 936.00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-589-870B-2 (1-183) x US-08-831-399-1 (1-638)

```
QY 1 MetArgLysIleValAlaAlaIleAlaValSerLeuThrThrValSerIleThrAla 20
DB 50 ATGCGCAGATCGTGTGACACCTTCGCGGTTTCCCTGACACCGGTCTGATACGGCC 109
QY 21 SerAlaSerAlaAspProSerIleAspSerIleValSerAlaAlaGluAlaGly 40
DB 110 AGCGCTTGGCAGACCTCCCAAGACTCGAAGGCCCAAGTCTCGCGCGCCAGGCCGAC 169
QY 41 IleThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAlaAsp 60
DB 170 ATCACCAGCAGCTGTGACACAGCTCGGCTGACCTTCTGATCGGCGGCCGCGAC 229
QY 61 GlyAlaLeuThrGlyThrTyrGluSerAlaValGlyAsnAlaGluSerArgTyrValLeu 80
DB 230 GGCGCCCTGACCGGAACCTACAGAGTCGGCCCTTCGCGCAAGCCGCTACGCTCTG 289
QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTrpThr 100
DB 290 ACCGCTGTGACAGAGCGGCCCGCCAGCCAGCGGACCGGACCGCCCTCGGTGAGC 349
QY 101 ValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyr 120
DB 350 GTGGCTTGGAGAAATTAACCTACCGCAACGCCCACTCCGACACCACTGAGCGGCCAGTAC 409
QY 121 ValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuThrSerGlyThrThrGlu 140
DB 410 GTCCGCGGCGCGCAGCGGATCAACACCGAGTGGCTGCTGACCTCCGCGACCAACCGAG 469
QY 141 AlaAsnAlaTrpLysSerThrIleuValGlyHisAspThrPheThrLysValLysProSer 160
DB 470 GCCAAGCGCTGGAATTCACAGCTGTGCGCCACGACACCTTCAACCAAGGTGAAGCGCTCC 529
```

Qy	Db	Qy	Db
16	530	181	590
Alalalaseerlllekpmlaalalyalyvalagl	GCCCCCTTCATTCACGCGGGGGAAGAGCCGCGTCAACACGAGGACCCCTCGACGCC	ValGlnGln	GTTTCAGCAG
YVlaaenAGClYAspProlenAspAla		183	598

US-09-589-870B-2 (1-183) x US-09-366-862-1 (1-638)

QY 1 MetArgLysIleValAlaIleAlaIleValSerLeuThrThrValSerIleThrAla 20
DB 50 ATGCCAGATGCTGCTTCCAGCCATCGCCGTTTCCCTTACCAACGAGCTCGATTACGGCC 109

QY 21 SerAlaSerAlaAspProSerIleAspSerIleAspSerIleValSerAlaAlaGluAlaGly 40
DB 110 AGCGCTTCGGCAGACCCCTCCAGAGCCTCGAAGGCCAGAGCTCGGCCGCCGAGGCCGCGC 169

QY 41 IleThrGlyThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 60
DB 170 ATCACCAGGACCTGCTACCAACAGCTCGGCTCGACCTTATGTCGACCGCGGCGCGAC 229

QY 61 GlAlaLeuThrGlyThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 80
DB 230 GCGCGCTTCAGCGAAGCCTACAGAGTGGCGCTCGGCAACGCGAGAGCGCGTACGTCCTG 289

QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
DB 290 ACCGGTCGTTACGACAGCGCCCGGCCACGACGAGCGGACCGCCCTCGATTGAGACG 349

QY 101 ValAlaIleTyrPlyAsnAsnTyrArgAsnAlaHisSerAlaThrThrThrThrThrThr 120
DB 350 GTGGCTTGAGAAATPACTACCGCAACGCGCACTCGGACCAACGCGAGCGCGCAGTAC 409

QY 121 ValGlyGlyAlaGluAlaArgIleAsnThrGlnTyrPleuLeuThrSerGlyThrThrGlu 140
DB 410 GTCCGCGCGCGCGAGCGAGATCAACACCCAGTGGCTGTACTCTCGGACACCGACG 469

QY 141 AlaAsnAlaTyrPlySerThrLeuValGlyHisAspThrPheThrIleValIleProSer 160
DB 470 GCCAAGCGCTGAGAAATPACTACCGCTGTGCGGCCACGACACCTTCACCAAGGTGAAGCGTCC 529

QY 161 AlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsnGlyAsnProLeuAspAla 180
DB 530 GCCGCTTCATGACGCGCGGAGAAAGCGCGGTCAACACGCAACCGCGCTCGACGCGC 589

QY 181 ValGlnGln 183
DB 590 GTTCAGCAG 598

RESULT 5

US-09-368-772-1
; Sequence 1, Application US/09368772
; Patent No. 6417331
; GENERAL INFORMATION:
; APPLICANT: Kopeckí, Erhard; Müller, Rainer;
; APPLICANT: Engh, Richard; Schmitt, Urban; Deger, Arno; Brandstetter, Hans
; TITLE OF INVENTION: Recombinant Inactive Core Streptavidin Mutants
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/368,772
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/831,399
; FILING DATE: 1-April-1997
; APPLICATION NUMBER: DE 196 13 053.0
; FILING DATE: 1-April-1996
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: DE 196 37 718.8
FILING DATE: 16-September-1996
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6417331man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: HUBR 1105
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEO ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 638 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 50..598
OTHER INFORMATION: Positions 50..121 correspond to sig peptide, and
OTHER INFORMATION: 122..598 to mat peptide.
US-09-368-772-1

Alignment Scores:
Pred. No.: 3,696-91 Length: 638
Score: 936..00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-09-589-870B-2 (1-183) x US-09-368-772-1 (1-638)

QY 1 MetArgLysIleValAlaIleAlaIleValSerLeuThrThrValSerIleThrAla 20
DB 50 ATGCCAGATGCTGCTTCCAGCCATCGCCGTTTCCCTTACCAACGAGCTCGATTACGGCC 109

QY 21 SerAlaSerAlaAspProSerIleAspSerIleAspSerIleValSerAlaAlaGluAlaGly 40
DB 110 AGCGCTTCGGCAGACCCCTCCAGAGACTCGAAGGCCAGCTCGGCCGCCGAGCGCGC 169

QY 41 IleThrGlyThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 60
DB 170 ATCACCAGGACCTGCTACCAACAGCTCGGCTCGACCTTATGTCGACCGCGGCGCGAC 229

QY 61 GlAlaLeuThrGlyThrThrThrThrThrThrThrThrThrThrThrThrThrThrThrThr 80
DB 230 GCGCGCTTCAGCGAAGCCTACAGAGTGGCGCTCGGCAACGCGAGAGCGCGTACGTCCTG 289

QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
DB 290 ACCGGTCGTTACGACAGCGCCCGGCCACGACGCGACCGCGCCCTCGATTGAGACG 349

QY 101 ValAlaIleTyrPlyAsnAsnTyrArgAsnAlaHisSerAlaThrThrThrThrThrThr 120
DB 350 GTGGCTTGAGAAATPACTACCGCAACGCGCACTCGGACCAACGCGAGCGCGCAGTAC 409

QY 121 ValGlyGlyAlaGluAlaArgIleAsnThrGlnTyrPleuLeuThrSerGlyThrThrGlu 140
DB 410 GTCCGCGCGCGCGAGCGAGATCAACACCCAGTGGCTGTACTCTCGGACACCGACG 469

QY 141 AlaAsnAlaTyrPlySerThrLeuValGlyHisAspThrPheThrIleValIleProSer 160
DB 470 GCCAAGCGCTGAGAAATPACTACCGCTGTGCGGCCACGACACCTTCACCAAGGTGAAGCGTCC 529

QY 161 AlaAlaSerIleAspAlaAlaLysLysAlaGlyValAsnAsnGlyAsnProLeuAspAla 180
DB 530 GCCGCTTCATGACGCGCGGAGAAAGCGCGGTCAACACGCAACCGCGCTCGACGCGC 589

QY 181 ValGlnGln 183
DB 590 GTTCAGCAG 598

RESULT 6

5168049-1
; Patent No. 5168049
; APPLICANT: MEADE, HARRY M.; GARWIN, JEFFREY L.
; TITLE OF INVENTION: PRODUCTION OF STREPTAVIDIN-LIKE
; POLYPEPTIDES
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/185,329
; FILING DATE: 21-APR-1988
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 656,873
; FILING DATE: 02-OCT-1984
; SEQ ID NO: 1:
; LENGTH: 1131
5168049-1

Alignment Scores:
Pred. No.: 8,176-91 Length: 1131
Score: 936.00 Matches: 183
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-09-589-870B-2 (1-183) x 5168049-1 (1-1131)

QY 1 MetArgLysGleValValAlaAlaIleAlaValSerLeuThrThyValSerIleThrAla 20
DB 480 ATGGCGAAGTCTCTGTCGAGCATCGCCGTTCCCTGCACCGCTCTGATTACGGCC 539
QY 21 SerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGluAlaGly 40
DB 540 AGGCTTCGGACACCCCTCTCAAGACTCGAAGGCCAGCTTCGGCGCCGACGGCCGG 599
QY 41 IleThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAlaAsp 60
DB 600 ATCACCAGGACCTGGTATCAACACAGCTCGGCTGACCTTATGTGACCGCGGGCCGAC 659
QY 61 GlyAlaLeuThrGlyThrTyrGluSerAlaValAlaGlyAsnAlaGluSerArgTyrValLeu 80
DB 660 GGGGCGCTGACCGAAGCTACAGAGTCGGCGTGGGACCGCCAGAGCGCTACGTCCTG 719
QY 81 ThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrThr 100
DB 720 ACCGCTCGTACACAGCGCCCGCCGACCGACGGCGGCGCCCTCGCTGGTGGAGCG 779
QY 101 ValAlaIleTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGlnTyr 120
DB 780 GTGGCTGGAGAAATTAATACCGCAACGCCCACTCCGCGACACGTCGAGCGCCAGTAC 839
QY 121 ValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThrGlu 140
DB 840 GTCGGGGGGCCGAGCGAGGATCAACACCACTGCTGCTGCTCCGCGACACCGAG 899
QY 141 AlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValLysProSer 160
DB 900 GCCAAGCCTGGAAAGTCCAGCTGCTCGGCCACGACACTTCCACCAAGGTGAAGCCGTC 959
QY 161 AlaAlaSerIleAsnAlaAlaLysLysAlaGlyValAsnAsnGlyAsnProLeuAspAla 180
DB 960 GCCGCTCCATCGACCGGGGAAGAGCGCGGCTCAACACGGCAACCGCTCGACGCGC 1019
QY 181 ValGlnGln 183
DB 1020 GTTCAGCAG 1028

RESULT 7
US-08-491-988-4
; Sequence 4, Application US/08491988
; Patent No. 597316
; GENERAL INFORMATION:
; APPLICANT: EPENETOS, AGAMENON A.
; APPLICANT: SPOONER, ROBERT A.

APPLICANT: DEONARAIN, MAHENDRA
TITLE OF INVENTION: Compounds for targeting
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: MAULAY NISSEN GOLDBERG KIEL & HAND, LLP
STREET: 261 MADISON AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016-2391
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/491,988
FILING DATE: 18-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: GOLDBERG, JULES E.
REGISTRATION NUMBER: 24,408
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-986-4090
TELEFAX: 212-818-9479
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1356 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 40..1344
US-08-491-988-4

Alignment Scores:
Pred. No.: 4,176-81 Length: 1356
Score: 846.00 Matches: 162
Percent Similarity: 99.39% Conservative: 1
Best Local Similarity: 98.78% Mismatches: 1
Query Match: 90.38% Indels: 0
DB: Gaps: 0

US-09-589-870B-2 (1-183) x US-08-491-988-4 (1-1356)

QY 20 AlaSerAlaSerAlaAspProSerLysAspSerLysAlaGlnValSerAlaAlaGluAla 39
DB 853 GCTGGCGGACCTGCGAGACCCGTCGAAGACTCCAAAGCTCAGGTTTCTGACCGCAAGCT 912
QY 40 GlyIleThrGlyThrTrpTyrAsnGlnLeuGlySerThrPheIleValThrAlaGlyAla 59
DB 913 GGTATACACGCGACCTCGTATACCACTCGGGGTGCACTTATCTGACCCCTGGTGG 972
QY 60 AspGlyAlaLeuThrGlyThrTyrGluSerAlaValAlaGlyAsnAlaGluSerArgTyrVal 79
DB 973 GACGAGGCTCTGACCTGCGCACCTAGCAATCTCGGTGGTGAACGAAATCCCGCTACGTA 1032
QY 80 LeuThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAlaLeuGlyTyrP 99
DB 1033 CTGACTGGCGCTTATGACTCTGACCTGCGCACGATGCTGTGTAACCGCTCGGGCTGG 1092
QY 100 ThrValAlaTrpLysAsnAsnTyrArgAsnAlaHisSerAlaThrThrTrpSerGlyGln 119
DB 1093 ACTGTGGCTGGAAAAACAATATGTAATGCGACAGCGCACTACGTGCTGGCCAA 1152
QY 120 TyrValGlyGlyAlaGluAlaArgIleAsnThrGlnTrpLeuLeuThrSerGlyThrThr 139
DB 1153 TACGTTGGCGGTGCTGAGCTCGTATCAACACTGAGCTGTAAATCCCGCACTACC 1212
QY 140 GluAlaAsnAlaTrpLysSerThrLeuValGlyHisAspThrPheThrLysValLysPro 159

Db 1213 GAAGGATGATGGAATCGACCTAGTAGTCACTGACACCTTACCAAGTAAAGCT 1272
QY 160 SerAlaAlaSerIleAspAlaAlaValysAlaGlyValAsnAsnGlyAsnProleuasp 179
Db 1273 TCTGCTGATGATTGATGCTGCCAAGAACGAGCGTAAACACCGTAAACCTCTAGAC 1332
QY 180 AlaValGlnGln 183
Db 1333 GCTGTTCAAGCAA 1344
RESULT 8
US-07-924-028A-2
; Sequence 2, Application US/07924028A
; Patent No. 5470573
; GENERAL INFORMATION:
; APPLICANT: Lubitz, Werner, Szostak, Michael P.
; TITLE OF INVENTION: CARRIER-BOUND RECOMBINANT PROTEINS, PROCESS
; TITLE OF INVENTION: FOR THE PRODUCTION AND USE AS IMMUNOGENS AND VACCINES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/924,028A
; FILING DATE: 30-SEP-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP91/00308
; FILING DATE: 02-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE 40 05 874
; FILING DATE: 24-FEB-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5470573man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: HUBR 1027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 525 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-07-924-028A-2
Alignment Scores:
Pred. No.: 9.01e-81 Length: 525
Score: 837.50 Matches: 164
Percent Similarity: 95.93% Conservative: 1
Best Local Similarity: 95.35% Mismatches: 6
Query Match: 89.48% Indels: 1
DB: 1 Gaps: 1
US-09-589-870B-2 (1-183) x US-07-924-028A-2 (1-525)
QY 12 SerLeuThrThrValSerIleThrAlaSerAlaAspProSerIysAspSerIys 31
Db 1 TCTAGAACTAGTAGATCCATC--GAGGCTAGGTCTATGACCCGCTCAAGACTCCAAA 57
QY 32 AlaGlnValSerAlaAlaGlnValGlyIleThrGlyThrTyrAsnGlnLeuGlySer 51

Db 58 GCTCAGTTTCTGACAGCCGAACTGATATCACTGGCACCTGGTATACCAACTGGGGTGC 117
QY 52 ThrPheIleValIleThrAlaGlyAlaAspGlyAlaLeuThrGlyThrTyrGlySerAlaVal 71
Db 118 ACTTTCATTGTGACCGCTGCTGGCGGACCGAGCTGATGCTGACCTAGCAATCTCGGTT 177
QY 72 GlyAsnAlaGlySerArgTyrValLeuThrGlyArgTyrAspSerAlaProAlaThrAsp 91
Db 178 GGTAAACGAGATCCCGCTAGCTACTGACTGGCCGTATGACTGTGACACTGGCCACCAT 237
QY 92 GlySerGlyThrAlaLeuGlyTyrThrValAlaTatPlyAsnAsnTyrArgAsnAlaHis 111
Db 238 GGCTCTGTACCGCTCTGGGCTGACCTGTGCTGGAACCACTATGTAATCGGAC 297
QY 112 SerAlaThrThrTyrSerGlyGlnTyrValGlyGlyAlaGlnAlaArgIleAsnThrGln 131
Db 298 ACGCCCACTAGCTGTGCTGGCCAAACGTTGGCGGTGCTGAGGCTCGATACACTCAG 357
QY 132 TyrLeuLeuThrSerGlyThrThrGlnAlaAsnAlaTyrPlySerThrLeuValGlyHis 151
Db 358 TGGCTGTAAACATCCGCACTAACCGAAGCATGCAATCGACACTAGTAGTGCAT 417
QY 152 AspThrPheThrIleValysValysProSerAlaIleSerIleAspAlaAlaValysAlaGly 171
Db 418 GACACCTTTACCAAGTTAAGCTTCTGCTGTACATTGATGCTGCCAAGAAACAGGC 477
QY 172 ValAsnAsnGlyAsnProLeuAspAlaValGlnGln 183
Db 478 GTAAACAAACGTAACCTCTAGACGCTGTTACGCAA 513
RESULT 9
US-08-491-988-6
; Sequence 6, Application US/08491988
; Patent No. 5973116
; GENERAL INFORMATION:
; APPLICANT: EPERNETOS, AGAMENON A.
; APPLICANT: SPOONER, ROBERT A.
; APPLICANT: DEONARAIN, MAHENDRA
; TITLE OF INVENTION: Compounds for targeting
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MCANILAY NISSEN GOLDBERG KIEL & HAND, LLP
; STREET: 261 MADISON AVENUE
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10016-2391
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/491,988
; FILING DATE: 18-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: GOLDBERG, JULES E.
; REGISTRATION NUMBER: 24,408
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-986-4090
; TELEFAX: 212-818-9479
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1296 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 40..1284

```

US-08-491-988-6

Alignment Scores:
Pred. No.: 2,328-70 Length: 1296
Score: 745.00 Matches: 142
Percent Similarity: 99.31% Conservative: 1
Best local Similarity: 98.61% Mismatches: 1
Query Match: 79.59% Indels: 0
DB: 2 Gaps: 0

US-09-589-870B-2 (1-183) x US-08-491-988-6 (1-1296)

QY 20 AAsESeAlAsESeAlAsPProSeSerYAsPSeSerYAsAlaGlnValSeSeAlAlaGlnAla 39
DB 853 GCTGGCGGACCTCGACGACCCGCTCAAGAGCTCAAGGCTCAAGTTCTGACGCGAAGCT 912
QY 40 GYlIeThrGlyThrTrpTYrAnGlnIeuGlySeSerThrPheIleValThrAlaGlyAla 59
DB 913 GGTATCACTGGGACCTGGATATACCAACTGGGGGTGACTTTTCATTGTACCGCTGGTGG 972
QY 60 AepGlyAlaLeuThrGlyThrTYrGlnSeSerAlaValGlyAsnAlaGlnSeArGTYrVal 79
DB 973 GACGAGCGCTGACCTGCGACCCACCTACCAATCTGGGTTGGTAAACGCAAGAAATCCGCTACCTTA 1032
QY 80 LeuThrGlyAsGTYrAsPSeSerAlaProAlaThrAsPGLySeSerGlyThrAlaIeuGlyTrp 99
DB 1033 CTGACTGGCGCTGATGACTCTGACACTGCGACCGAGTGGCTGTGGTACCGCTCTGGGCTAGG 1092
QY 100 ThrValAlaTrpLYrAsnAsnTYrArGAsnAlaHisSerAlaThrThrTrpSerGlyGln 119
DB 1093 ACTGGGCTTGGAAAAACACATGCTGTAATGGGACACAGCGCACATACGTTGGTGGCCAA 1152
QY 120 TYrValGlyGlyAlaGlnAlaArgIleAsnThrGlnTrpLeuThrSeSerGlyThrThr 139
DB 1153 TACGTTGGCGGGCTGAGAGCTCTGTTCAACACTCGTGGCTTAAACATCCGGGCACTTACC 1212
QY 140 GlnAlaAsnAlaTrpLYrSeSerThrIeuValGlyHisAsPThrPheThrIlyeValIysPro 159
DB 1213 GAAGGAAATGCATGAGAAATCCACACTAGTAGTCTGACACCTTAAACAAAGTTAAGCT 1272
QY 160 SerAlaAlaSer 163
DB 1273 TCTGCTGCTAGC 1284

RESULT 10
US-09-142-974B-4
; Sequence 4, Application US/09142974B
; Patent No. 6451995
; GENERAL INFORMATION:
; APPLICANT: Cheung, Nai-Kong V.
; APPLICANT: Larson, Steven M.
; APPLICANT: Guo, Hong-Fen
; APPLICANT: Rivlin, Ken
; TITLE OF INVENTION: Single Chain FV Constructs of Anti-Ganglioside GD2
; TITLE OF INVENTION: Antibodies
; FILE REFERENCE: MSK.P-013-USNP
; CURRENT APPLICATION NUMBER: US/09/142,974B
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: PCT/US97/04427
; PRIOR FILING DATE: 1997-03-20
; PRIOR APPLICATION NUMBER: 60/013,703
; PRIOR FILING DATE: 1996-03-20
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1173
; TYPE: DNA
; ORGANISM: Murine
; FEATURE:
; OTHER INFORMATION: 366-bcfv-sctseptavidin
US-09-142-974B-4

```

Alignment Scores:	Pred. No.:	2.74e-64	Length:	1173
Score:	687.50	Matches:	133	
Percent Similarity:	92.52%	Conservative:	4	
Best Local Similarity:	90.48%	Mismatches:	3	
Query Match:	73.45%	Indels:	7	
DB:	4	Gaps:	1	

US-09-589-8708-2 (1-183) x US-09-142-974B-4 (1-1173)

QY	17	SeqIleThraIaSerAlaSerAlaAspProSerLyAspSerLyAlaGlnValSerAla	36
DB	697	TCAgTCACCGCTCTCTCTACGGCC-----GCTGAATCCGGTGCT	735
QY	37	AlaGlnAlaGlyIleThrGlyThrTyrAsnGlnLeuGlySerThrPheIleValThr	56
DB	736	GCTGAACAGAGTATCACCGGACCTGTGTACACAGCTCGGCTGACTTCATCGTGACC	795
QY	57	AlaGlyAlaAspGlyAlaLeuThrGlyThrTyrGlnSerAlaValGlyAsnAlaGlnSer	76
DB	796	GCGGGCCGCCACCGCGCCCTGACCGGAACCTACGAGTCGCGCCGCGCAACGCGGAGAGC	855
QY	77	ArgTyrAlaLeuThrGlyArgTyrAspSerAlaProAlaThrAspGlySerGlyThrAla	96
DB	856	CGCTACGCTCTGACCGGTCGTTACGACAGCGCCCGGCCACCGAGCGGACGACCGCC	915
QY	97	LeuGlyTrpThrValAlaTyrPLeuAsnAsnTyrArgAsnAlaHisSerAlaThrThrTyr	116
DB	916	CTCGGTTGGACCGTGCGCTCGAAGAAATPACTACCGCAACCCCACTCCGGACCACTGG	975
QY	117	SerGlyGlnTyrValGlyGlyAlaGlnAlaArgIleAsnThrGlnTyrPLeuLeuThrSer	136
DB	976	AGCGGCCAGTACGTCGCGCGCGCGGACGAGCGAGCATCAACCCCGTGGCTGTGACCTCC	1035
QY	137	GlyThrThrGlnAlaAsnAlaTyrPLeuSerThrLeuValGlyHisAspThrPheThrLys	156
DB	1036	GGCAACACCGAGGCCACGCTGTGAGATCCACGCTGTGTGGCCAGCACACTTCACCAAG	1095
QY	157	ValIysProSerAlaAlaSer	163
DB	1096	GTGAAGCCGTCTCGCGCCCTCC	1116

RESULT 11
US-09-142-974B-3
Sequence 3, Application US/09142974B
Patent No. 6451995
GENERAL INFORMATION:
APPLICANT: Cheung, Nai-Kong V.
APPLICANT: Larson, Steven M.
APPLICANT: Guo, Hong-Fen
APPLICANT: Rivlin, Ken
APPLICANT: Sadelaia, Michel
TITLE OF INVENTION: Single Chain FV Constructs of Anti-Ganglioside GD2
FILE REFERENCE: MSK-P-013-USNP
CURRENT APPLICATION NUMBER: US/09/142,974B
CURRENT FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: PCT/US97/04427
PRIOR FILING DATE: 1997-03-20
PRIOR APPLICATION NUMBER: 60/013,703
PRIOR FILING DATE: 1996-03-20
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 3
LENGTH: 1176
TYPE: DNA
ORGANISM: Murine
FEATURE:
OTHER INFORMATION: 5f11-scFv-streptavidin
NAME/KEY: unsure
LOCATION: (37)
NAME/KEY: unsure
LOCATION: (79)

